## Preface

- Audience ........................................... xix
- Documentation Accessibility ............................ xix
- Related Documents .................................. xx
- Conventions ...................................... xx

## What's New in Oracle Text? 

- Oracle Database 11g Release 2 (11.2) New Features in Oracle Text........ xxv
- Oracle Database 11g Release 1 (11.1) New Features in Oracle Text........ xxvi

### 1 Oracle Text SQL Statements and Operators

- ALTER INDEX ............................................................ 1-2
- ALTER TABLE: Supported Partitioning Statements .................. 1-16
- CATSEARCH .................................................. 1-21
- CONTAINS .............................................................. 1-28
- CREATE INDEX ...................................................... 1-36
- DROP INDEX ......................................................... 1-58
- MATCHES ............................................................ 1-59
- MATCH_SCORE ...................................................... 1-61
- SCORE ........................................................................ 1-62

### 2 Oracle Text Indexing Elements

- Overview .......................................................... 2-1
  - Creating Preferences .............................................. 2-2
- datastore Types .................................................. 2-2
  - DIRECT_DATASTORE ...................................... 2-3
    - DIRECT_DATASTORE CLOB Example .................. 2-3
  - MULTI_COLUMN_DATASTORE .......................... 2-3
  - Indexing and DML ................................................. 2-4
  - MULTI_COLUMN_DATASTORE Restriction .............. 2-4
  - MULTI_COLUMN_DATASTORE Example ................. 2-4
  - MULTI_COLUMN_DATASTORE Filter Example ........... 2-4
  - Tagging Behavior ................................................. 2-5
  - Indexing Columns as Sections .............................. 2-5
- DETAIL_DATASTORE .......................................... 2-6
  - Synchronizing Master/Detail Indexes .................... 2-7
  - Example Master/Detail Tables ............................. 2-7
    - Master Table Example ........................................ 2-7
    - Detail Table Example ......................................... 2-7
    - Detail Table Example Attributes ....................... 2-7
    - Master/Detail Index Example .............................. 2-8
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FILE_DATASTORE</td>
<td>2-8</td>
</tr>
<tr>
<td>PATH Attribute Limitations</td>
<td>2-9</td>
</tr>
<tr>
<td>FILE_DATASTORE and Security</td>
<td>2-9</td>
</tr>
<tr>
<td>FILE_DATASTORE Example</td>
<td>2-10</td>
</tr>
<tr>
<td>URL_DATASTORE</td>
<td>2-10</td>
</tr>
<tr>
<td>URL Syntax</td>
<td>2-10</td>
</tr>
<tr>
<td>URL_DATASTORE Attributes</td>
<td>2-11</td>
</tr>
<tr>
<td>URL_DATASTORE and Security</td>
<td>2-12</td>
</tr>
<tr>
<td>URL_DATASTORE Example</td>
<td>2-12</td>
</tr>
<tr>
<td>USER_DATASTORE</td>
<td>2-12</td>
</tr>
<tr>
<td>Constraints</td>
<td>2-13</td>
</tr>
<tr>
<td>Editing Procedure after Indexing</td>
<td>2-13</td>
</tr>
<tr>
<td>USER_DATASTORE with CLOB Example</td>
<td>2-13</td>
</tr>
<tr>
<td>USER_DATASTORE with BLOB_LOC Example</td>
<td>2-14</td>
</tr>
<tr>
<td>NESTED_DATASTORE</td>
<td>2-14</td>
</tr>
<tr>
<td>NESTED_DATASTORE Example</td>
<td>2-15</td>
</tr>
<tr>
<td>Create the Nested Table</td>
<td>2-15</td>
</tr>
<tr>
<td>Insert Values into Nested Table</td>
<td>2-15</td>
</tr>
<tr>
<td>Create Nested Table Preferences</td>
<td>2-16</td>
</tr>
<tr>
<td>Create Index on Nested Table</td>
<td>2-16</td>
</tr>
<tr>
<td>Query Nested Datstore</td>
<td>2-16</td>
</tr>
<tr>
<td>Filter Types</td>
<td>2-16</td>
</tr>
<tr>
<td>CHARSET_FILTER</td>
<td>2-17</td>
</tr>
<tr>
<td>UTF-16 Big- and Little-Endian Detection</td>
<td>2-18</td>
</tr>
<tr>
<td>Indexing Mixed-Character Set Columns</td>
<td>2-18</td>
</tr>
<tr>
<td>Indexing Mixed-Character Set Example</td>
<td>2-18</td>
</tr>
<tr>
<td>AUTO_FILTER</td>
<td>2-19</td>
</tr>
<tr>
<td>Indexing Formatted Documents</td>
<td>2-20</td>
</tr>
<tr>
<td>Explicitly Bypassing Plain Text or HTML in Mixed Format Columns</td>
<td>2-20</td>
</tr>
<tr>
<td>Character Set Conversion With AUTO_FILTER</td>
<td>2-21</td>
</tr>
<tr>
<td>NULL_FILTER</td>
<td>2-21</td>
</tr>
<tr>
<td>Indexing HTML Documents</td>
<td>2-21</td>
</tr>
<tr>
<td>MAIL_FILTER</td>
<td>2-21</td>
</tr>
<tr>
<td>Filter Behavior</td>
<td>2-22</td>
</tr>
<tr>
<td>About the Mail Filter Configuration File</td>
<td>2-23</td>
</tr>
<tr>
<td>Mail File Configuration File Structure</td>
<td>2-23</td>
</tr>
<tr>
<td>Mail_Filter Example</td>
<td>2-24</td>
</tr>
<tr>
<td>USER_FILTER</td>
<td>2-24</td>
</tr>
<tr>
<td>Using USER_FILTER with Charset and Format Columns</td>
<td>2-25</td>
</tr>
<tr>
<td>Explicitly Bypassing Plain Text or HTML in Mixed Format Columns</td>
<td>2-25</td>
</tr>
<tr>
<td>Character Set Conversion with USER_FILTER</td>
<td>2-26</td>
</tr>
<tr>
<td>User Filter Example</td>
<td>2-26</td>
</tr>
<tr>
<td>PROCEDURE_FILTER</td>
<td>2-27</td>
</tr>
<tr>
<td>Parameter Order</td>
<td>2-29</td>
</tr>
<tr>
<td>Procedure Filter Execute Requirements</td>
<td>2-29</td>
</tr>
<tr>
<td>Error Handling</td>
<td>2-29</td>
</tr>
<tr>
<td>Procedure Filter Preference Example</td>
<td>2-30</td>
</tr>
</tbody>
</table>
Lexer Types

AUTO_LEXER
  AUTO_LEXER Attributes Inherited from BASIC_LEXER
  AUTO_LEXER Language-Independent Attributes
  AUTO_LEXER Language-Dependent Attributes
  AUTO_LEXER User-Defined Dictionary Attributes

BASIC_LEXER
  Stemming User-Dictionaries
  BASIC_LEXER Example

MULTI_LEXER
  Multi-language Stoplists
  MULTI_LEXER Example
  Querying Multi-Language Tables

CHINESE_VGRAM_LEXER
  CHINESE_VGRAM_LEXER Attribute
  Character Sets

CHINESE_LEXER
  CHINESE_LEXER Attribute
  Customizing the Chinese Lexicon

JAPANESE_VGRAM_LEXER
  JAPANESE_VGRAM_LEXER Attributes
  JAPANESE_VGRAM_LEXER Character Sets

JAPANESE_LEXER
  Customizing the Japanese Lexicon
  JAPANESE_LEXER Attributes
  JAPANESE_LEXER Character Sets
  Japanese Lexer Example

KOREAN_MORPH_LEXER
  Supplied Dictionaries
  Supported Character Sets
  Unicode Support
    Limitations on Korean Unicode Support
  KOREAN_MORPH_LEXER Attributes
  Limitations
  KOREAN_MORPH_LEXER Example: Setting Composite Attribute
    NGRAM Example
    COMPONENT_WORD Example

USER_LEXER
  Limitations
  USER_LEXER Attributes

INDEX_PROCEDURE
  Requirements
  Parameters
  Restrictions

INPUT_TYPE
  VARCHAR2 Interface
  CLOB Interface
3 Oracle Text CONTAINS Query Operators

Operator Precedence .............................................................................................................. 3-2
Group 1 Operators .................................................................................................................. 3-2
Group 2 Operators and Characters ....................................................................................... 3-2
Procedural Operators ............................................................................................................. 3-3
Precedence Examples ............................................................................................................. 3-3
Altering Precedence ................................................................................................................ 3-3
ABOUT .................................................................................................................................. 3-4
ACCUMulate (, ) ..................................................................................................................... 3-7
<table>
<thead>
<tr>
<th>Feature</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>AND (&amp;)</td>
<td>3-9</td>
</tr>
<tr>
<td>Broader Term (BT, BTG, BTP, BTI)</td>
<td>3-10</td>
</tr>
<tr>
<td>DEFINEMERGE</td>
<td>3-12</td>
</tr>
<tr>
<td>DEFINESCORE</td>
<td>3-13</td>
</tr>
<tr>
<td>EQUIVAlence (=)</td>
<td>3-17</td>
</tr>
<tr>
<td>Fuzzy</td>
<td>3-18</td>
</tr>
<tr>
<td>HASPATH</td>
<td>3-20</td>
</tr>
<tr>
<td>INPATH</td>
<td>3-22</td>
</tr>
<tr>
<td>MDATA</td>
<td>3-28</td>
</tr>
<tr>
<td>MINUS (-)</td>
<td>3-30</td>
</tr>
<tr>
<td>MNOT</td>
<td>3-31</td>
</tr>
<tr>
<td>Narrower Term (NT, NTG, NTP, NTI)</td>
<td>3-32</td>
</tr>
<tr>
<td>NDATA</td>
<td>3-34</td>
</tr>
<tr>
<td>NEAR ()</td>
<td>3-36</td>
</tr>
<tr>
<td>NOT (-)</td>
<td>3-40</td>
</tr>
<tr>
<td>OR (|)</td>
<td>3-41</td>
</tr>
<tr>
<td>Preferred Term (PT)</td>
<td>3-42</td>
</tr>
<tr>
<td>Related Term (RT)</td>
<td>3-43</td>
</tr>
<tr>
<td>SDATA</td>
<td>3-44</td>
</tr>
<tr>
<td>soundex (!)</td>
<td>3-47</td>
</tr>
<tr>
<td>stem ($)</td>
<td>3-48</td>
</tr>
<tr>
<td>Stored Query Expression (SQE)</td>
<td>3-49</td>
</tr>
<tr>
<td>SYNonym (SYN)</td>
<td>3-50</td>
</tr>
<tr>
<td>threshold (&gt;)</td>
<td>3-51</td>
</tr>
<tr>
<td>Translation Term (TR)</td>
<td>3-52</td>
</tr>
<tr>
<td>Translation Term Synonym (TRSYN)</td>
<td>3-53</td>
</tr>
<tr>
<td>Top Term (TT)</td>
<td>3-55</td>
</tr>
<tr>
<td>weight (*)</td>
<td>3-56</td>
</tr>
<tr>
<td>wildcards (%__)</td>
<td>3-58</td>
</tr>
<tr>
<td>WITHIN</td>
<td>3-60</td>
</tr>
</tbody>
</table>

### 4 Special Characters in Oracle Text Queries

- Grouping Characters ........................................ 4-1
- Escape Characters ........................................ 4-1
  - Querying Escape Characters ............................ 4-2
- Reserved Words and Characters ....................... 4-2

### 5 CTX_ADM Package

- MARK_FAILED .............................................. 5-2
- RECOVER ................................................... 5-3
- SET_PARAMETER ........................................... 5-4

### 6 CTX_CLS Package

- TRAIN ...................................................... 6-2
- CLUSTERING ............................................... 6-5
7 CTX_DDL Package

ADD_ATTR_SECTION ................................................................. 7-3
ADD_FIELD_SECTION .............................................................. 7-4
ADD_INDEX ......................................................................... 7-7
ADD_MDATA ........................................................................ 7-9
ADD_MDATA_COLUMN ........................................................... 7-11
ADD_MDATA_SECTION ........................................................... 7-12
ADD_NDATA_SECTION ............................................................ 7-13
ADD_SDATA_COLUMN ............................................................. 7-14
ADD_SDATA_SECTION ............................................................ 7-16
ADD_SPECIAL_SECTION ....................................................... 7-18
ADD_STOPCLASS ................................................................. 7-20
ADD_STOPSECTION ............................................................... 7-21
ADD_STOPTHEME ................................................................. 7-23
ADD_STOPWORD ................................................................ 7-24
ADD_SUB_LEXER ................................................................. 7-26
ADD_ZONE_SECTION ............................................................. 7-28
COPY_POLICY ...................................................................... 7-31
CREATE_INDEX_SET .............................................................. 7-32
CREATE_POLICY ................................................................ 7-33
CREATE_PREFERENCE .......................................................... 7-35
CREATE_SECTION_GROUP ....................................................... 7-38
CREATE_SHADOW_INDEX ....................................................... 7-41
CREATE_STOPLIST ............................................................... 7-43
DROP_INDEX_SET ................................................................. 7-45
DROP_POLICY ..................................................................... 7-46
DROP_PREFERENCE .............................................................. 7-47
DROP_SECTION_GROUP .......................................................... 7-48
DROP_SHADOW_INDEX ........................................................... 7-49
DROP_STOPLIST ................................................................. 7-50
EXCHANGE_SHADOW_INDEX .................................................. 7-51
OPTIMIZE_INDEX ................................................................. 7-53
POPULATE_PENDING .......................................................... 7-58
RECREATE_INDEX_ONLINE ...................................................... 7-59
REMOVE_INDEX ................................................................ 7-66
REMOVE_MDATA ................................................................. 7-67
REMOVE_SECTION .............................................................. 7-68
REMOVE_STOPCLASS ........................................................... 7-69
REMOVE_STOPTHEME ........................................................... 7-70
REMOVE_STOPWORD ............................................................ 7-71
REMOVE_SUB_LEXER ............................................................ 7-72
REPLACE_INDEX_METADATA .................................................. 7-73
SET_ATTRIBUTE .................................................................. 7-74
SYNC_INDEX ...................................................................... 7-75
UNSET_ATTRIBUTE ............................................................. 7-78
UPDATE_POLICY .................................................................. 7-79
### 8  CTX_DOC Package

<table>
<thead>
<tr>
<th>Function</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FILTER</td>
<td>8-3</td>
</tr>
<tr>
<td>GIST</td>
<td>8-6</td>
</tr>
<tr>
<td>HIGHLIGHT</td>
<td>8-10</td>
</tr>
<tr>
<td>IFILTER</td>
<td>8-13</td>
</tr>
<tr>
<td>MARKUP</td>
<td>8-14</td>
</tr>
<tr>
<td>PKENCODE</td>
<td>8-19</td>
</tr>
<tr>
<td>POLICY_FILTER</td>
<td>8-20</td>
</tr>
<tr>
<td>POLICY_GIST</td>
<td>8-21</td>
</tr>
<tr>
<td>POLICY_HIGHLIGHT</td>
<td>8-24</td>
</tr>
<tr>
<td>POLICY_MARKUP</td>
<td>8-26</td>
</tr>
<tr>
<td>POLICY_SNIPPET</td>
<td>8-29</td>
</tr>
<tr>
<td>POLICY_THEMES</td>
<td>8-31</td>
</tr>
<tr>
<td>POLICY_TOKENS</td>
<td>8-33</td>
</tr>
<tr>
<td>SET_KEY_TYPE</td>
<td>8-35</td>
</tr>
<tr>
<td>SNIPPET</td>
<td>8-36</td>
</tr>
<tr>
<td>THEMES</td>
<td>8-39</td>
</tr>
<tr>
<td>TOKENS</td>
<td>8-42</td>
</tr>
</tbody>
</table>

### 9  CTX_ENTITY Package

<table>
<thead>
<tr>
<th>Function</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD_EXTRACT_RULE</td>
<td>9-2</td>
</tr>
<tr>
<td>ADD_STOP_ENTITY</td>
<td>9-5</td>
</tr>
<tr>
<td>COMPILE</td>
<td>9-6</td>
</tr>
<tr>
<td>CREATE_EXTRACT_POLICY</td>
<td>9-8</td>
</tr>
<tr>
<td>DROP_EXTRACT_POLICY</td>
<td>9-9</td>
</tr>
<tr>
<td>EXTRACT</td>
<td>9-10</td>
</tr>
<tr>
<td>REMOVE_EXTRACT_RULE</td>
<td>9-12</td>
</tr>
<tr>
<td>REMOVE_STOP_ENTITY</td>
<td>9-13</td>
</tr>
</tbody>
</table>

### 10  CTX_OUTPUT Package

<table>
<thead>
<tr>
<th>Function</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD_EVENT</td>
<td>10-2</td>
</tr>
<tr>
<td>ADD_TRACE</td>
<td>10-3</td>
</tr>
<tr>
<td>DISABLE_QUERY_STATS</td>
<td>10-5</td>
</tr>
<tr>
<td>ENABLE_QUERY_STATS</td>
<td>10-6</td>
</tr>
<tr>
<td>END_LOG</td>
<td>10-7</td>
</tr>
<tr>
<td>END_QUERY_LOG</td>
<td>10-8</td>
</tr>
<tr>
<td>GET_TRACE_VALUE</td>
<td>10-9</td>
</tr>
<tr>
<td>LOG_TRACES</td>
<td>10-10</td>
</tr>
<tr>
<td>LOGFILENAME</td>
<td>10-11</td>
</tr>
<tr>
<td>REMOVE_EVENT</td>
<td>10-12</td>
</tr>
<tr>
<td>REMOVE_TRACE</td>
<td>10-13</td>
</tr>
<tr>
<td>RESET_TRACE</td>
<td>10-14</td>
</tr>
<tr>
<td>START_LOG</td>
<td>10-15</td>
</tr>
<tr>
<td>START_QUERY_LOG</td>
<td>10-16</td>
</tr>
</tbody>
</table>
11  CTX_QUERY Package
BROWSE_WORDS........................................................................................................ 11-2
COUNT_HITS............................................................................................................... 11-5
EXPLAIN..................................................................................................................... 11-6
HFEEDBACK ............................................................................................................... 11-9
REMOVE_SQE........................................................................................................... 11-13
RESULT_SET.............................................................................................................. 11-14
STORE_SQE............................................................................................................... 11-20

12  CTXREPORT Package
Procedures in CTX_REPORT..................................................................................... 12-1
Using the Function Versions.................................................................................... 12-1
DESCRIBE_INDEX...................................................................................................... 12-3
DESCRIBE_POLICY................................................................................................... 12-4
CREATE_INDEX_SCRIPT............................................................................................ 12-5
CREATE_POLICY_SCRIPT.......................................................................................... 12-6
INDEX_SIZE........................................................................................................... 12-7
INDEX_STATS........................................................................................................... 12-8
QUERY_LOG_SUMMARY............................................................................................ 12-12
TOKEN_INFO............................................................................................................ 12-16
TOKEN_TYPE............................................................................................................ 12-18

13  CTX_THES Package
ALTER_PHRASE........................................................................................................ 13-3
ALTER_THESAURUS.................................................................................................. 13-5
BT............................................................................................................................... 13-6
BTG............................................................................................................................. 13-8
BTI............................................................................................................................... 13-10
BTP............................................................................................................................. 13-12
CREATE_PHRASE..................................................................................................... 13-14
CREATE_RELATION.................................................................................................. 13-15
CREATE_THESAURUS............................................................................................... 13-17
CREATE_TRANSLATION............................................................................................. 13-18
DROP_PHRASE........................................................................................................ 13-19
DROP_RELATION...................................................................................................... 13-20
DROP_THESAURUS.................................................................................................... 13-22
DROP_TRANSLATION............................................................................................... 13-23
HAS_RELATION........................................................................................................ 13-24
NT............................................................................................................................... 13-25
NTG............................................................................................................................. 13-27
NTI............................................................................................................................... 13-29
NTP............................................................................................................................. 13-31
OUTPUT_STYLE....................................................................................................... 13-33
PT............................................................................................................................... 13-34
RT............................................................................................................................... 13-36
SN............................................................................................................................... 13-38
14  CTX_ULEXER Package

WILDCARD_TAB................................................................. 14-2

15  Oracle Text Utilities

Thesaurus Loader (ctxload) .............................................................. 15-1
  Text Loading ........................................................................ 15-1
  ctxload Syntax.................................................................... 15-1
    Mandatory Arguments...................................................... 15-2
    Optional Arguments......................................................... 15-2
  ctxload Examples............................................................. 15-3
    Thesaurus Import Example .............................................. 15-3
    Thesaurus Export Example .............................................. 15-3
Entity Extraction User Dictionary Loader (ctxload) ........................................ 15-3
  ctxload Syntax.................................................................... 15-3
    Mandatory Arguments...................................................... 15-3
    Considerations When Creating a User Dictionary ............... 15-4
    XML Schema ..................................................................... 15-4
  ctxload Example ................................................................. 15-5
Knowledge Base Extension Compiler (ctxkbtc) ............................................. 15-6
  Knowledge Base Character Set.............................................. 15-6
  ctxkbtc Syntax.................................................................... 15-6
  ctxkbtc Usage Notes.......................................................... 15-7
  ctxkbtc Limitations............................................................ 15-7
  ctxkbtc Constraints on Thesaurus Terms .............................. 15-7
  ctxkbtc Constraints on Thesaurus Relations ............................ 15-8
  Extending the Knowledge Base ............................................ 15-8
    Example for Extending the Knowledge Base ...................... 15-9
  Adding a Language-Specific Knowledge Base ...................... 15-9
    Limitations for Adding a Knowledge Base ......................... 15-10
  Order of Precedence for Multiple Thesauri ......................... 15-10
  Size Limits for Extended Knowledge Base .......................... 15-10
Lexical Compiler (ctxlc) .................................................................... 15-10
  Syntax of ctxlc .................................................................... 15-11
    Mandatory Arguments...................................................... 15-11
    Optional Arguments........................................................ 15-11
  Performance Considerations .............................................. 15-11
  ctxlc Usage Notes .............................................................. 15-11
  Example ............................................................................... 15-12
16 Oracle Text Alternative Spelling

Overview of Alternative Spelling Features ................................................................. 16-1
Alternate Spelling ...................................................................................................... 16-2
Base-Letter Conversion ............................................................................................. 16-2
  Generic Versus Language-Specific Base-Letter Conversions ............................ 16-2
New German Spelling ............................................................................................... 16-2

Overriding Alternative Spelling Features ................................................................. 16-3
Overriding Base-Letter Transformations with Alternate Spelling ......................... 16-3

Alternative Spelling Conventions ............................................................................ 16-3
  German Alternate Spelling Conventions ............................................................... 16-4
  Danish Alternate Spelling Conventions ............................................................... 16-4
  Swedish Alternate Spelling Conventions ............................................................. 16-4

A Oracle Text Result Tables

CTX_QUERY Result Tables ......................................................................................... A-1
EXPLAIN Table ......................................................................................................... A-1
  Operation Column Values ................................................................................. A-2
  OPTIONS Column Values .............................................................................. A-2
HFEEDBACK Table ................................................................................................. A-3
  Operation Column Values ............................................................................. A-4
  OPTIONS Column Values ............................................................................. A-4
  CTX_FEEDBACK_TYPE .................................................................................. A-4

CTX_DOC Result Tables ......................................................................................... A-5
  Filter Table ....................................................................................................... A-5
  Gist Table ........................................................................................................ A-6
  Highlight Table ................................................................................................ A-6
  Markup Table .................................................................................................... A-6
  Theme Table ..................................................................................................... A-7
  Token Table ...................................................................................................... A-7

CTX_THES Result Tables and Data Types ............................................................ A-7
  EXP_TAB Table Type ....................................................................................... A-8

B Oracle Text Supported Document Formats

About Document Filtering Technology ................................................................... B-1
  Latest Updates for Patch Releases ..................................................................... B-1
  Restrictions on Format Support ......................................................................... B-1
  Supported Platforms for AUTO_FILTER Document Filtering Technology .......... B-2
    Supported Platforms ..................................................................................... B-2
  Filtering on PDF Documents and Security Settings ......................................... B-2
  PDF Filtering Limitations ............................................................................... B-3
  Environment Variables .................................................................................... B-3
  General Limitations ........................................................................................ B-3

Supported Document Formats ................................................................................. B-4
  Word Processing and Desktop Publishing Formats ......................................... B-4
  Spreadsheet Formats ...................................................................................... B-6
  Presentation Formats ....................................................................................... B-6
Knowledge Base ............................................................................................................................... D-7
Knowledge Base Extension .............................................................................................................. D-7
Multilingual Features Matrix ........................................................................................................... D-7

E Oracle Text Supplied Stoplists

- English Default Stoplist................................................................................................................. E-1
- Chinese Stoplist (Traditional) ........................................................................................................ E-2
- Chinese Stoplist (Simplified) ......................................................................................................... E-2
- Danish (dk) Default Stoplist ......................................................................................................... E-3
- Dutch (nl) Default Stoplist ........................................................................................................... E-3
- Finnish (sf) Default Stoplist ......................................................................................................... E-4
- French (f) Default Stoplist ............................................................................................................ E-5
- German (d) Default Stoplist ......................................................................................................... E-6
- Italian (i) Default Stoplist ............................................................................................................. E-7
- Portuguese (pt) Default Stoplist ................................................................................................ E-7
- Spanish (e) Default Stoplist ........................................................................................................ E-7
- Swedish (s) Default Stoplist ......................................................................................................... E-8

F The Oracle Text Scoring Algorithm

- Scoring Algorithm for Word Queries .......................................................................................... F-1
  - Word Scoring Example .............................................................................................................. F-2
  - DML and Scoring Algorithm .................................................................................................. F-2

G Oracle Text Views

- CTX_CLASSES ............................................................................................................................... G-2
- CTX_EXTRACT_POLICIES .............................................................................................................. G-3
- CTX_EXTRACT_POLICY_VALUES ............................................................................................... G-3
- CTX_FILTER_BY_COLUMNS .......................................................................................................... G-3
- CTX_INDEXES ............................................................................................................................... G-4
- CTX_INDEX_ERRORS .................................................................................................................... G-4
- CTX_INDEX_OBJECTS ................................................................................................................... G-5
- CTX_INDEX_PARTITIONS ............................................................................................................. G-5
- CTX_INDEX_SETS ........................................................................................................................ G-5
- CTX_INDEX_SET_INDEXES .......................................................................................................... G-6
- CTX_INDEX_SUB_LEXERS .......................................................................................................... G-6
- CTX_INDEX_SUB_LEXER_VALUES .............................................................................................. G-6
- CTX_INDEX_VALUES ................................................................................................................... G-6
- CTX_OBJECTS ............................................................................................................................. G-7
- CTX_OBJECT_ATTRIBUTES ........................................................................................................... G-7
- CTX_OBJECT_ATTRIBUTE_LOV .................................................................................................... G-7
- CTX_ORDER_BY_COLUMNS .......................................................................................................... G-8
- CTX_PARAMETERS ...................................................................................................................... G-8
- CTX_PENDING ............................................................................................................................ G-9
- CTX_PREFERENCES ..................................................................................................................... G-10
- CTX_PREFERENCE_VALUES ....................................................................................................... G-10
- CTX_SECTIONS ........................................................................................................................... G-10
Stopword Transformations in Oracle Text

Understanding Stopword Transformations ................................................................. H-1
  Word Transformations ......................................................................................... H-2
  AND Transformations ......................................................................................... H-2
  OR Transformations ............................................................................................. H-2
  ACCUMulate Transformations .............................................................................. H-3
  MINUS Transformations ....................................................................................... H-3
  MNOT Transformations ......................................................................................... H-3
  NOT Transformations ............................................................................................ H-3
  EQUIVAlence Transformations .............................................................................. H-4
  NEAR Transformations .......................................................................................... H-4
Weight Transformations .............................................................................................................. H-5
Threshold Transformations ....................................................................................................... H-5
WITHIN Transformations ....................................................................................................... H-5

1 AUTO_LEXER Parts-of-Speech Tagging

Tagging in Arabic ..................................................................................................................... I-1
Tagging in Catalan .................................................................................................................... I-3
Tagging in Chinese - Traditional and Simplified ................................................................. I-6
Tagging in Croatian ................................................................................................................ I-6
Tagging in Danish ..................................................................................................................... I-7
Tagging in Dutch ....................................................................................................................... I-10
Tagging in English .................................................................................................................... I-12
Tagging in Farsi ........................................................................................................................ I-14
Tagging in Finnish ..................................................................................................................... I-17
Tagging in French ....................................................................................................................... I-19
Tagging in German ...................................................................................................................... I-21
Tagging in Italian ........................................................................................................................ I-24
Tagging in Japanese .................................................................................................................... I-27
Tagging in Korean ....................................................................................................................... I-28
Tagging in Bokmal ....................................................................................................................... I-30
Tagging in Nynorsk ..................................................................................................................... I-33
Tagging in Portuguese ............................................................................................................... I-36
Tagging in Russian ...................................................................................................................... I-39
Tagging in Slovak ......................................................................................................................... I-41
Tagging in Slovenian ................................................................................................................. I-42
Tagging in Spanish ....................................................................................................................... I-44
Tagging in Swedish ..................................................................................................................... I-46

Index
This Preface contains these topics:

- Audience
- Documentation Accessibility
- Related Documents
- Conventions

Audience

Oracle Text Reference is intended for Oracle Text application developers or system administrators who perform the following task:

- maintain an Oracle Text system

To use this document, you need experience with the Oracle relational database management system, SQL, SQL*Plus, and PL/SQL.

Documentation Accessibility

Our goal is to make Oracle products, services, and supporting documentation accessible, to all users, including users that are disabled. To that end, our documentation includes features that make information available to users of assistive technology. This documentation is available in HTML format, and contains markup to facilitate access by the disabled community. Accessibility standards will continue to evolve over time, and Oracle is actively engaged with other market-leading technology vendors to address technical obstacles so that our documentation can be accessible to all of our customers. For more information, visit the Oracle Accessibility Program Web site at

http://www.oracle.com/accessibility/

Accessibility of Code Examples in Documentation

Screen readers may not always correctly read the code examples in this document. The conventions for writing code require that closing braces should appear on an otherwise empty line; however, some screen readers may not always read a line of text that consists solely of a bracket or brace.
Accessibility of Links to External Web Sites in Documentation

This documentation may contain links to Web sites of other companies or organizations that Oracle does not own or control. Oracle neither evaluates nor makes any representations regarding the accessibility of these Web sites.

Access to Oracle Support

Oracle customers have access to electronic support through My Oracle Support. For information, visit http://www.oracle.com/support/contact.html or visit http://www.oracle.com/accessibility/support.html if you are hearing impaired.

Related Documents

For more information about Oracle Text, refer to:

- Oracle Text Application Developer’s Guide

For more information about Oracle Database, refer to:

- Oracle Database Concepts
- Oracle Database Administrator’s Guide
- Oracle Database Utilities
- Oracle Database Performance Tuning Guide
- Oracle Database SQL Reference
- Oracle Database Reference
- Oracle Database Application Developer’s Guide - Fundamentals

For more information about PL/SQL, refer to:

- PL/SQL User’s Guide and Reference

Conventions

This section describes the conventions used in the text and code examples of this documentation set. It describes:

- Conventions in Text
- Conventions in Code Examples
- Conventions for Windows Operating Systems

Conventions in Text

We use various conventions in text to help you more quickly identify special terms. The following table describes those conventions and provides examples of their use.

<table>
<thead>
<tr>
<th>Convention</th>
<th>Meaning</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bold</td>
<td>Bold typeface indicates terms that are defined in the text or terms that appear in a glossary, or both.</td>
<td>When you specify this clause, you create an index-organized table.</td>
</tr>
<tr>
<td>Italic</td>
<td>Italic typeface indicates book titles or emphasis.</td>
<td>Oracle Database Concepts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ensure that the recovery catalog and target database do not reside on the same disk.</td>
</tr>
</tbody>
</table>
Conventions in Code Examples

Code examples illustrate SQL, PL/SQL, SQL*Plus, or other command-line statements. They are displayed in a monospace (fixed-width) font and separated from normal text as shown in this example:

```sql
SELECT username FROM dba_users WHERE username = 'MIGRATE';
```

The following table describes typographic conventions used in code examples and provides examples of their use.

<table>
<thead>
<tr>
<th>Convention</th>
<th>Meaning</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UPPERCASE</strong></td>
<td>Uppercase monospace typeface indicates elements supplied by the system. Such elements include parameters, privileges, datatypes, Recovery Manager keywords, SQL keywords, SQL*Plus or utility commands, packages and methods, as well as system-supplied column names, database objects and structures, usernames, and roles.</td>
<td>You can specify this clause only for a NUMBER column.</td>
</tr>
<tr>
<td><strong>lowercase</strong></td>
<td>Lowercase monospace typeface indicates executable programs, filenames, directory names, and sample user-supplied elements. Such elements include computer and database names, net service names and connect identifiers, user-supplied database objects and structures, column names, packages and classes, usernames and roles, program units, and parameter values. Note: Some programmatic elements use a mixture of <strong>UPPERCASE</strong> and lowercase. Enter these elements as shown.</td>
<td>Enter sqlplus to start SQL*Plus. The password is specified in the orapwd file. Back up the datafiles and control files in the /disk1/oracle/dbs directory. The department_id, department_name, and location_id columns are in the hr.departments table. Set the QUERY_REWRITE_ENABLED initialization parameter to true. Connect as oe user. The JRepUtil class implements these methods.</td>
</tr>
<tr>
<td><strong>lowercase italic</strong></td>
<td>Lowercase italic monospace font represents placeholders or variables.</td>
<td>You can specify the <code>parallel_clause</code>. Run <code>old_release.SQL</code> where <code>old_release</code> refers to the release you installed prior to upgrading.</td>
</tr>
</tbody>
</table>

**Conventions in Code Examples**

Code examples illustrate SQL, PL/SQL, SQL*Plus, or other command-line statements. They are displayed in a monospace (fixed-width) font and separated from normal text as shown in this example:

```sql
SELECT username FROM dba_users WHERE username = 'MIGRATE';
```

The following table describes typographic conventions used in code examples and provides examples of their use.

<table>
<thead>
<tr>
<th>Convention</th>
<th>Meaning</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ]</td>
<td>Anything enclosed in brackets is optional.</td>
<td>DECIMAL (digits [ , precision ])</td>
</tr>
<tr>
<td>{ }</td>
<td>Braces are used for grouping items.</td>
<td>(ENABLE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ENABLE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[COMPRESS</td>
</tr>
<tr>
<td>...</td>
<td>Ellipsis points mean repetition in syntax descriptions.</td>
<td>CREATE TABLE ... AS subquery;</td>
</tr>
<tr>
<td></td>
<td>In addition, ellipsis points can mean an omission in code examples or text.</td>
<td>SELECT col1, col2, ..., coln FROM employees;</td>
</tr>
<tr>
<td>Other symbols</td>
<td>You must use symbols other than brackets ([ ]), braces ({ }), vertical bars (</td>
<td>), and ellipsis points (...) exactly as shown.</td>
</tr>
<tr>
<td><strong>Italics</strong></td>
<td>Italicized text indicates placeholders or variables for which you must supply particular values.</td>
<td>CONNECT SYSTEM/system_password DB_NAME = database_name</td>
</tr>
</tbody>
</table>
The following table describes conventions for Windows operating systems and provides examples of their use.

<table>
<thead>
<tr>
<th>Convention</th>
<th>Meaning</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPPERCASE</td>
<td>Uppercase typeface indicates elements supplied by the system. We show these terms in uppercase in order to distinguish them from terms you define. Unless terms appear in brackets, enter them in the order and with the spelling shown. Because these terms are not case sensitive, you can use them in either UPPERCASE or lowercase.</td>
<td>SELECT last_name, employee_id FROM employees; SELECT * FROM USER_TABLES; DROP TABLE hr.employees;</td>
</tr>
<tr>
<td>lowercase</td>
<td>Lowercase typeface indicates user-defined programmatic elements, such as names of tables, columns, or files. Note: Some programmatic elements use a mixture of UPPERCASE and lowercase. Enter these elements as shown.</td>
<td>SELECT last_name, employee_id FROM employees; sqlplus hr/hr CREATE USER mjones IDENTIFIED BY ty3MU9;</td>
</tr>
</tbody>
</table>

**Conventions for Windows Operating Systems**

The following table describes conventions for Windows operating systems and provides examples of their use.

<table>
<thead>
<tr>
<th>Convention</th>
<th>Meaning</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choose Start &gt; menu item</td>
<td>How to start a program.</td>
<td>To start the Database Configuration Assistant, choose Start &gt; Programs &gt; Oracle - HOME_NAME &gt; Configuration and Migration Tools &gt; Database Configuration Assistant.</td>
</tr>
<tr>
<td>File and directory names</td>
<td>File and directory names are not case sensitive. The following special characters are not allowed: left angle bracket (&lt;), right angle bracket (&gt;), colon (:), double quotation marks (&quot;), slash (/), pipe (</td>
<td>), and dash (-). The special character backslash () is treated as an element separator, even when it appears in quotes. If the filename begins with , then Windows assumes it uses the Universal Naming Convention.</td>
</tr>
<tr>
<td>C:&gt;</td>
<td>Represents the Windows command prompt of the current hard disk drive. The escape character in a command prompt is the caret (^). Your prompt reflects the subdirectory in which you are working. Referred to as the command prompt in this manual.</td>
<td>C:\oracle\oradata&gt;</td>
</tr>
<tr>
<td>Special characters</td>
<td>The backslash () special character is sometimes required as an escape character for the double quotation mark (&quot;) special character at the Windows command prompt. Parentheses and the single quotation mark (’) do not require an escape character. Refer to your Windows operating system documentation for more information on escape and special characters.</td>
<td>C:&gt; exp HR/HR TABLES=emp QUERY=&quot;WHERE job='REP'&quot;</td>
</tr>
<tr>
<td>HOME_NAME</td>
<td>Represents the Oracle home name. The home name can be up to 16 alphanumeric characters. The only special character allowed in the home name is the underscore.</td>
<td>C:&gt; net start OracleHOME_NAME\TNSListener</td>
</tr>
</tbody>
</table>
In releases prior to Oracle8i release 8.1.3, when you installed Oracle components, all subdirectories were located under a top level `ORACLE_HOME` directory. The default for Windows NT was `C:\orant`.

This release complies with Optimal Flexible Architecture (OFA) guidelines. All subdirectories are not under a top level `ORACLE_HOME` directory. There is a top level directory called `ORACLE_BASE` that by default is `C:\oracle\product\10.1.0`. If you install the latest Oracle release on a computer with no other Oracle software installed, then the default setting for the first Oracle home directory is `C:\oracle\product\10.1.0\db_n`, where `n` is the latest Oracle home number. The Oracle home directory is located directly under `ORACLE_BASE`.

All directory path examples in this guide follow OFA conventions.

Refer to Oracle Database Installation Guide for Windows for additional information about OFA compliances and for information about installing Oracle products in non-OFA compliant directories.

### Table

<table>
<thead>
<tr>
<th>Convention</th>
<th>Meaning</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>ORACLE_HOME</code> and <code>ORACLE_BASE</code></td>
<td>In releases prior to Oracle8i release 8.1.3, when you installed Oracle components, all subdirectories were located under a top level <code>ORACLE_HOME</code> directory. The default for Windows NT was <code>C:\orant</code>. This release complies with Optimal Flexible Architecture (OFA) guidelines. All subdirectories are not under a top level <code>ORACLE_HOME</code> directory. There is a top level directory called <code>ORACLE_BASE</code> that by default is <code>C:\oracle\product\10.1.0</code>. If you install the latest Oracle release on a computer with no other Oracle software installed, then the default setting for the first Oracle home directory is <code>C:\oracle\product\10.1.0\db_n</code>, where <code>n</code> is the latest Oracle home number. The Oracle home directory is located directly under <code>ORACLE_BASE</code>.</td>
<td>Go to the <code>ORACLE_BASE\ORACLE_HOME\rdbms\admin</code> directory.</td>
</tr>
</tbody>
</table>
What's New in Oracle Text?

The following describes new features of the Oracle Database 11g Release 2 (11.2) edition of Oracle Text and provides pointers to additional information. Information about new features from previous releases is also retained to help you migrate to the current release.

The following sections describe the new features in Oracle Text:

- Oracle Database 11g Release 2 (11.2) New Features in Oracle Text
- Oracle Database 11g Release 1 (11.1) New Features in Oracle Text

Oracle Database 11g Release 2 (11.2) New Features in Oracle Text

- Entity extraction and identification
  You can search for terms that are unknown to you without specifying a particular search text. This process of identifying names, places, dates, and other objects when they are mentioned in a document and tagging each occurrence with its type and subtype correctly is called entity extraction.

  The value of entity extraction is that it enables you to identify instances of a particular pre-specified class of entities in textual documents, thus enabling you to produce a structured view of a document that can later be used for text/data mining and more comprehensive intelligence analysis.

  See Also: Oracle Text Application Developer’s Guide for further information

- Name matching
  Someone accustomed to the spelling rules of one culture can have difficulty applying those same rules to a name originating from a different culture. Name matching provides a solution to match proper names that might differ in spelling due to orthographic variation. It also enables you to search for somewhat inaccurate data, such as might occur when a record’s first name and surname are not properly segmented.

  See Also: Oracle Text Application Developer’s Guide for further information

- Result set interface
  A page of search results typically consists of many disparate elements, such as metadata of the first few documents, per-word hit counts, or total hit counts. In past releases of Oracle Text, generating these results required several queries and calls, such as a query on the base table, a call to `CTX_QUERY.COUNT_HITS`, and so on. Each call required time to reparse the query and look up index metadata. In this release, instead of accessing the database to construct bits of the search results, you can use the result set interface, which is able to produce the various kinds of data needed for a page of search results all at once, thus improving performance by sharing overhead. The result set interface can also return data views that are difficult to express in SQL, such as top N by category queries.
Oracle Database 11g Release 1 (11.1) New Features in Oracle Text

- On Windows systems, the executable file that you specify for the USER_FILTER command attribute must now exist in the %ORACLE_HOME%/ctx/bin directory instead of %ORACLE_HOME%/bin.

  See Also:  USER_FILTER on page 2-24

- Zero downtime for applications with new incremental indexing and online index creation.

  See Also:  "Creating a CONTEXT Index Incrementally with POPULATE_PENDING" in Oracle Text Application Developer’s Guide

- New features for re-creating an index online and finer control for maintenance processes.

  See Also:  "Re-Creating an Index” in Oracle Text Application Developer’s Guide

- New Oracle Text Manager in Oracle Enterprise Manager with which you can:
  - Monitor health of Oracle Text indexes.
  - Modify index settings.
  - Generate index-level statistics about disk space, fragmentation, garbage, frequency of words, and more.
  - Synchronize, optimize, and rebuild indexes.
  - Diagnose problems, and resume failed operations.
  - Manage logs.

  See Also:  "Text Manager in Oracle Enterprise Manager” in Oracle Text Application Developer’s Guide

- New support for composite domain index for CONTEXT indextype for improved mixed-query performance.

  See Also:  "Composite Domain Index (CDI) in Oracle Text’ in Oracle Text Application Developer’s Guide

- Improved query performance and scalability.

  See Also:  "Parallelizing Queries Across Oracle RAC Nodes” in Oracle Text Application Developer’s Guide

- New SDATA section type and SDATA operator that enable range searches on metadata.

  See Also:  "SDATA Section” in Oracle Text Application Developer’s Guide
- New user-defined scoring feature, `DEFINESCORE` and `DEFINEMERGE`.

  **See Also:**  "DEFINESCORE" on page 3-13 and "DEFINEMERGE" on page 3-12

- New `AUTO_LEXER` lexer type that performs language identification, word segmentation, document analysis, part-of-speech tagging, and stemming. The `AUTO_LEXER` type also enables customization of these components.

  **See Also:**  "AUTO_LEXER" on page 2-31

- New values for the `INDEX_STEMS` attribute of the `BASIC_LEXER` type to enable better query performance for stem ($) queries.

  **See Also:**  "BASIC_LEXER" on page 2-39

- `NOPOPULATE` option for `ALTER INDEX` to support incremental indexing.

  **See Also:**  "POPULATE | NOPOPULATE" on page 1-46

- New limit for the number of partitions in Oracle Text is now the same as the maximum for Oracle Database.

  **See Also:**  "Partitioned Tables and Indexes" in Oracle Text Application Developer’s Guide

- New usage tracking feature.

  **See Also:**  "Database Feature Usage Tracking in Oracle Enterprise Manager" in Oracle Text Application Developer’s Guide
This chapter describes the SQL statements and Oracle Text operators for creating and managing Oracle Text indexes and performing Oracle Text queries.

The following statements are described in this chapter:

- ALTER INDEX
- ALTER TABLE: Supported Partitioning Statements
- CATSEARCH
- CONTAINS
- CREATE INDEX
- DROP INDEX
- MATCHES
- MATCH_SCORE
- SCORE
Purpose

Use ALTER INDEX to make changes to, or perform maintenance tasks for a CONTEXT, CTXCAT, or CTXRULE index.

All Index Types

Use ALTER INDEX to perform the following tasks on all Oracle Text index types:

- Rename the index or index partition. See "ALTER INDEX RENAME Syntax" on page 1-4.
- Rebuild the index using different preferences. Some restrictions apply for the CTXCAT index type. See "ALTER INDEX REBUILD Syntax" on page 1-4.
- Add stopwords to the index. See "ALTER INDEX REBUILD Syntax" on page 1-4.

CONTEXT and CTXRULE Index Types

Use ALTER INDEX to perform the following tasks on CONTEXT and CTXRULE index types:

- Resume a failed index operation (creation/optimization).
- Add sections and stop sections to the index.
- Replace index metadata.

See Also:  "ALTER INDEX REBUILD Syntax" on page 1-4 to learn more about performing these tasks

Overview of ALTER INDEX Syntax

The syntax for ALTER INDEX is fairly complex. The major divisions are covered in the following sections:

- "ALTER INDEX MODIFY PARTITION Syntax" on page 1-3—use this to modify an index partition's metadata.
- "ALTER INDEX PARAMETERS Syntax" on page 1-3—use this to modify the parameters of a nonpartitioned index, or to modify all partitions of a local partitioned index, without rebuilding the index.
- "ALTER INDEX RENAME Syntax" on page 1-4—use this to rename an index or index partition.
- "ALTER INDEX REBUILD Syntax" on page 1-4—use this to rebuild an index or index partition. With this statement, you can also replace index metadata; add stopwords, sections, and stop sections to an index; and resume a failed operation.

The parameters for ALTER INDEX REBUILD have their own syntax, which is a subset of the syntax for ALTER INDEX. For example, the ALTER INDEX REBUILD
ALTER INDEX PARAMETERS statement can take either REPLACE or RESUME as an argument, and ALTER INDEX REBUILD PARAMETERS (‘REPLACE’) can take several arguments. Valid examples of ALTER INDEX REBUILD include the following statements:

ALTER INDEX REBUILD PARALLEL n
ALTER INDEX REBUILD PARAMETERS (‘SYNC memsize’)  
ALTER INDEX REBUILD PARAMETERS (‘REPLACE DATASTORE datastore_pref’)  
ALTER INDEX REBUILD PARAMETERS (‘REPLACE WORDLIST wordlist_pref’)  

**ALTER INDEX MODIFY PARTITION Syntax**

Use the following syntax to modify the metadata of an index partition:

```
ALTER INDEX index_name MODIFY PARTITION partition_name PARAMETER (paramstring)
```

- **index_name**
  Specify the name of the index whose partition metadata you want to modify.

- **partition_name**
  Specify the name of the index partition whose metadata you want to modify.

- **paramstring**
  The only valid argument here is 'REPLACE METADATA'. This follows the same syntax as ALTER INDEX REBUILD PARTITION PARAMETERS (‘REPLACE METADATA’); see the REPLACE METADATA subsection of the "ALTER INDEX REBUILD Syntax" section on page 1-6 for more information. (The two statements are equivalent. ALTER INDEX MODIFY PARTITION is offered for ease of use, and is the recommended syntax.)

**ALTER INDEX PARAMETERS Syntax**

Use the following syntax to modify the parameters either of nonpartitioned or local partitioned indexes, without rebuilding the index. For partitioned indexes, this statement works at the index level, not at the partition level. This statement changes information for the entire index, including all partitions.

```
ALTER INDEX index_name PARAMETERS (paramstring)
```

- **paramstring**
  ALTER INDEX PARAMETERS accepts the following arguments for paramstring:

  - 'REPLACE METADATA'
    Replaces current metadata. See the REPLACE METADATA subsection of the "ALTER INDEX REBUILD Syntax" section on page 1-6 for more information.

  - 'ADD STOPWORD'
    Dynamically adds a stopword to an index. See the ADD STOPWORD subsection of the "ALTER INDEX REBUILD Syntax" section on page 1-10 for more information.

  - 'ADD FIELD SECTION'
    Dynamically adds a field section to an index. See the ADD FIELD subsection of the "ALTER INDEX REBUILD Syntax" section on page 1-10 for more information.

  - 'ADD ZONE SECTION'
    Dynamically adds a zone section to an index. See the ADD ZONE subsection of the "ALTER INDEX REBUILD Syntax" section on page 1-10 for more information.

  - 'ADD ATTR SECTION'
Dynamically adds an attribute section to an index. See the ADD ATTR subsection of the “ALTER INDEX REBUILD Syntax” section on page 1-11 for more information.

Each of the prior statements has an equivalent ALTER INDEX REBUILD PARAMETERS version. For example, ALTER INDEX PARAMETERS ('REPLACE METADATA') is equivalent to ALTER INDEX REBUILD PARAMETERS ('REPLACE METADATA'). However, the ALTER INDEX PARAMETERS versions work on either partitioned or nonpartitioned indexes, whereas the ALTER INDEX REBUILD PARAMETERS versions work only on nonpartitioned indexes.

**ALTER INDEX RENAME Syntax**

Use the following syntax to rename an index or index partition:

```
ALTER INDEX [schema.]index_name RENAME TO new_index_name;
ALTER INDEX [schema.]index_name RENAME PARTITION part_name TO new_part_name;
```

**[schema.]index_name**

Specify the name of the index to rename.

**new_index_name**

Specify the new name for [schema.]index. The new_index_name parameter can be no more than 25 bytes, and 21 bytes for a partitioned index. If you specify a name longer than 25 bytes (or longer than 21 bytes for a partitioned index), then Oracle Text returns an error and the renamed index is no longer valid.

---

**Note:** When new_index_name is more than 25 bytes (21 for local partitioned index) and less than 30 bytes, Oracle Text renames the index, even though the system returns an error. To drop the index and associated tables, you must drop new_index_name with the DROP INDEX statement and then re-create and drop index_name.

**part_name**

Specify the name of the index partition to rename.

**new_part_name**

Specify the new name for partition.

**ALTER INDEX REBUILD Syntax**

Use ALTER INDEX REBUILD to rebuild an index, rebuild an index partition, resume a failed operation, replace index metadata, add stopwords to an index, or add sections and stop sections to an index.

The ALTER INDEX REBUILD syntax has its own subsyntax. That is, its parameters have their own syntax. For example, the ALTER INDEX REBUILD PARAMETERS statement can take either REPLACE or RESUME as an argument, and ALTER INDEX REBUILD PARAMETERS ('REPLACE') has several arguments it can take.

Valid examples of ALTER INDEX REBUILD include the following statements:

```
ALTER INDEX REBUILD PARALLEL n
ALTER INDEX REBUILD PARAMETERS (SYNC memsize)
ALTER INDEX REBUILD PARAMETERS (REPLACE DATASTORE datastore_pref)
ALTER INDEX REBUILD PARAMETERS (REPLACE WORDLIST wordlist_pref)
```
This is the syntax for ALTER INDEX REBUILD:

```
ALTER INDEX [schema.]index [REBUILD] [PARTITION partname] [ONLINE] [PARAMETERS (paramstring)] [PARALLEL N];
```

**PARTITION partname**
Rebuilds the index partition `partname`. Only one index partition can be built at a time.

When you rebuild a partition you can specify only `RESUME` or `REPLACE` in `paramstring`. These operations work only on the `partname` you specify.

With the `REPLACE` operation, you can specify only `MEMORY` and `STORAGE` for each index partition.

**Adding Partitions**
To add a partition to the base table, use the `ALTER TABLE SQL` statement. When you add a partition to an indexed table, Oracle Text automatically creates the metadata for the new index partition. The new index partition has the same name as the new table partition. Change the index partition name with `ALTER INDEX RENAME`.

**Splitting or Merging Partitions**
Splitting or merging a table partition with `ALTER TABLE` renders the index partitions invalid. You must rebuild them with `ALTER INDEX REBUILD`.

**[ONLINE]**
Enables you to continue to perform updates, insertions, and deletions on a base table. It does not enable you to query the base table.

---

**Note:** You can specify `REPLACE` or `RESUME` when rebuilding an index or an index partition `ONLINE`.

---

**PARAMETERS (paramstring)**
Optionally specify `paramstring`. If you do not specify `paramstring`, then Oracle Text rebuilds the index with existing preference settings.

The syntax for `paramstring` is as follows:

```
paramstring =
  'REPLACE
   [DATASTORE datastore_pref]
   [FILTER filter_pref]
   [LEXER lexer_pref]
   [WORDLIST wordlist_pref]
   [STORAGE storage_pref]
   [STOPLIST stoplist]
   [SECTION GROUP section_group]
   [MEMORY memsize]
   [[POPULATE | NOPOPULATE]
    [INDEX SET index_set]]

  [METADATA preference new_preference]
  [METADATA FORMAT COLUMN format_column_name]
  [[METADATA] SYNC (MANUAL | EVERY "interval-string" | ON COMMIT)]
  [[METADATA] TRANSACTIONAL | NONTRANSACTIONAL]
  | RESUME [memory memsize]
  | OPTIMIZE [token index_token | fast | full [maxtime (time | unlimited)]
  | SYNC [memory memsize]
  | ADD STOPWORD word [language language]
```
ALTER INDEX

| ADD ZONE SECTION section_name tag tag |
| ADD FIELD SECTION section_name tag tag [(VISIBLE | INVISIBLE)] |
| ADD ATTR SECTION section_name tag tag@attr |
| ADD STOP SECTION tag' |

REPLACE [optional_preference_list]
Rebuilds an index. You can optionally specify your own preferences, or system-defined preferences.

You can replace only preferences that are supported for that index type. For instance, you cannot replace index set for a CONTEXT or CTXRULE index. Similarly, for the CTXCAT index type, you can replace lexer, wordlist, storage index set, and memory preferences.

The POPULATE parameter is the default and need not be specified. If you want to empty the index of its contents, then specify NOPOPULATE. Clear an index of its contents when you must rebuild your index incrementally. The NOPOPULATE choice is available for a specific partition of the index, and not just for the entire index.

If you are rebuilding a partitioned index using the REPLACE parameter, then you can specify only STORAGE, MEMORY, and NOPOPULATE.

See Also: Chapter 2, "Oracle Text Indexing Elements" for more information about creating and setting preferences, including information about system-defined preferences.

REPLACE METADATA preference new_preference
Replaces the existing preference class settings, including SYNC parameters, of the index with the settings from new_preference. Only index preferences and attributes are replaced. The index is not rebuilt.

This statement is useful for when you want to replace a preference and its attribute settings after the index is built, without reindexing all data. Reindexing data can require significant time and computing resources.

This statement is also useful for changing the SYNC parameter type, which can be automatic, manual, or on-commit.

The ALTER INDEX REBUILD PARAMETER ('REPLACE METADATA') statement does not work for a local partitioned index at the global level for the index. You cannot, for example, use this syntax to change a global preference, such as filter or lexer type, without rebuilding the index. Use ALTER INDEX PARAMETERS instead to change the metadata of an index at the global level, including all partitions. See "ALTER INDEX PARAMETERS Syntax" on page 1-3.

When should I use the METADATA keyword? REPLACE METADATA should be used only when the change in index metadata will not lead to an inconsistent index, which can lead to incorrect query results.

For example, use this statement in the following instances:

- To go from a single-language lexer to a multilexer in anticipation of multilingual data. For an example, see "Replacing Index Metadata: Changing Single-Lexer to Multilexer" on page 1-13.

- To change the WILDCARD_MAXTERMS setting in BASIC_WORDLIST.

- To change the SYNC parameter type, which can be automatic, manual, or on-commit.

These changes are safe and will not lead to an inconsistent index that might adversely affect your query results.
There can be many instances when changing metadata can result in inconsistent index data. For example, Oracle recommends against using the METADATA keyword after performing the following procedures:

- Changing the USER_DATASTORE procedure to a new PL/SQL stored procedure that has different output.
- Changing the BASIC_WORDLIST attribute PREFIX_INDEX from NO to YES because no prefixes have been generated for existing documents. Changing it from YES to NO is safe.
- Adding or changing BASIC_LEXER printjoin and skipjoin characters, because new queries with these characters would be lexed differently from how these characters were lexed at index time.

In these unsafe cases, Oracle recommends rebuilding the index.

**REPLACE [METADATA] SYNC (MANUAL | EVERY "interval-string" | ON COMMIT)**

Specifies SYNC for automatic synchronization of the CONTEXT index when a DML change has occurred to the base table. You can specify one of the SYNC methods shown in Table 1–1.

**Table 1–1  ALTER INDEX SYNC Methods**

<table>
<thead>
<tr>
<th>Sync Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MANUAL</td>
<td>No automatic synchronization. This is the default. You must manually synchronize the index using CTX_DDL.SYNC_INDEX. Use MANUAL to disable ON COMMIT and EVERY synchronization.</td>
</tr>
<tr>
<td>EVERY interval-string</td>
<td>Automatically synchronize the index at a regular interval specified by the value of interval-string. interval-string takes the same syntax as that for scheduler jobs. Automatic synchronization using EVERY requires that the index creator have CREATE JOB privileges. Make sure that interval-string is set to a long enough period so that any previous synchronization jobs will have completed. Otherwise, the sync job may hang. interval-string must be enclosed in double quotation marks (&quot; &quot;). See &quot;Enabling Automatic Index Synchronization&quot; on page 1-49 for an example of automatic sync syntax.</td>
</tr>
</tbody>
</table>
Each partition of a locally partitioned index can have its own type of sync: (ON COMMIT, EVERY, or MANUAL). The type of sync specified in master parameter strings applies to all index partitions unless a partition specifies its own type.

With automatic (EVERY) synchronization, you can specify memory size and parallel synchronization. The syntax is:

\[ \ldots \text{EVERY} \text{ interval_string} \text{ MEMORY} \text{ mem_size} \text{ PARALLEL} \text{ paradigm} \ldots \]

ON COMMIT synchronizations can only be executed serially and at the same memory size as what was specified at index creation.

**Note:** This command rebuilds the index. When you want to change the SYNC setting without rebuilding the index, use the `REBUILD REPLACE METADATA SYNC (MANUAL | ON COMMIT)` operation.

**REPLACE [METADATA] TRANSACTIONAL | NONTRANSACTIONAL**

This parameter enables you to turn the TRANSACTIONAL property on or off. For more information, see "TRANSACTIONAL" on page 1-47.

Using this parameter only succeeds if there are no rows in the DML pending queue. Therefore, you may need to sync the index before issuing this command.

To turn on the TRANSACTIONAL index property:

```
ALTER INDEX myidx REBUILD PARAMETERS('replace metadata transactional');
```

or

```
ALTER INDEX myidx REBUILD PARAMETERS('replace  transactional');
```

To turn off the TRANSACTIONAL index property:

```
ALTER INDEX myidx REBUILD PARAMETERS('replace metadata nontransactional');
```

or

```
ALTER INDEX myidx REBUILD PARAMETERS('replace  nontransactional');
```
RESUME [MEMORY memsize]
Resumes a failed index operation. You can optionally specify the amount of memory to use with memsize.

Note: This ALTER INDEX operation applies only to CONTEXT and CTXRULE indexes. It does not apply to CTXCAT indexes.

OPTIMIZE [token index_token | fast | full [maxtime (time | unlimited)]]

Optimizes the index. Specify token, fast, or full optimization. You typically optimize after you synchronize the index.

When you optimize in token mode, Oracle Text optimizes only index_token. Use this method of optimization to quickly optimize index information for specific words.

When you optimize in fast mode, Oracle Text works on the entire index, compacting fragmented rows. However, in fast mode, old data is not removed.

When you optimize in full mode, you can optimize the whole index or a portion. This method compacts rows and removes old data (deleted rows).

Note: Optimizing in full mode runs even when there are no deleted document rows. This is useful when you must optimize time-limited batches with the maxtime parameter.

Use the maxtime parameter to specify in minutes the time Oracle Text is to spend on the optimization operation. Oracle Text starts the optimization where it left off and optimizes until complete or until the time limit has been reached, whichever comes first. Specifying a time limit is useful for automating index optimization, where you set Oracle Text to optimize the index for a specified time on a regular basis.

When you specify maxtime unlimited, the entire index is optimized. This is the default. When you specify 0 for maxtime, Oracle Text performs minimal optimization.

Log the progress of optimization by writing periodic progress updates to the CTX_OUTPUT log. An event for CTX_OUTPUT_ADD_EVENT, called CTX_OUTPUT_EVENT_OPT_PRINT_TOKEN, prints each token as it is being optimized.

Note: This ALTER INDEX operation applies only to CONTEXT and CTXRULE indexes. It does not apply to CTXCAT indexes.

SYNC [MEMORY memsize]

Note: This ALTER INDEX operation will not be supported in future releases.

To synchronize your index, use CTX_DDL.SYNC_INDEX.
Synchronizes the index. You can optionally specify the amount of runtime memory to use with `memsize`. Synchronize the index when you have DML operations on your base table.

**Note:** This `ALTER INDEX` operation applies only to `CONTEXT` and `CTXRULE` indexes. It does not apply to `ctxcat` indexes.

**Memory Considerations** The memory parameter `memsize` specifies the amount of memory Oracle Text uses for the `ALTER INDEX` operation before flushing the index to disk. Specifying a large amount of memory improves indexing performance because there is less I/O and improves query performance and maintenance because there is less fragmentation.

Specifying smaller amounts of memory increases disk I/O and index fragmentation, but might be useful if you want to track indexing progress or when run-time memory is scarce.

**ADD STOPWORD** *word* *[language language]*

Dynamically adds a stopword `word` to the index.

Index entries for `word` that existed before this operation are not deleted. However, subsequent queries on `word` are treated as though it has always been a stopword.

When your stoplist is a multi-language stoplist, you must specify `language`.

The index is _not_ rebuilt by this statement.

**ADD ZONE SECTION** *section_name* *tag* *tag*

Dynamically adds the zone section `section_name` identified by `tag` to the existing index.

The added section `section_name` applies only to documents indexed after this operation. For the change to take effect, you must manually re-index any existing documents that contain the tag.

The index is _not_ rebuilt by this statement.

**Note:** This `ALTER INDEX` operation applies only to `CONTEXT` and `CTXRULE` indexes. It does not apply to `ctxcat` indexes.

**See Also:** "Notes" on page 1-15

**ADD FIELD SECTION** *section_name* *tag* *tag* *(VISIBLE | INVISIBLE)*

Dynamically adds the field section `section_name` identified by `tag` to the existing index.

Optionally specify `VISIBLE` to make the field sections visible. The default is `INVISIBLE`.

**See Also:** `CTX_DDL.ADD_FIELD_SECTION` for more information on visible and invisible field sections.

The added section `section_name` applies only to documents indexed after this operation. For the change to affect previously indexed documents, you must explicitly re-index the documents that contain the tag.

This statement does not rebuild the index.
ALTER INDEX

Note:  This ALTER INDEX operation applies only to CONTEXT CTXRULE indexes. It does not apply to CTXCAT indexes.

See Also:  "Notes" on page 1-15.

ADD ATTR SECTION section_name tag tag@attr
Dynamically adds an attribute section section_name to the existing index. You must specify the XML tag and attribute in the form tag@attr. You can add attribute sections only to XML section groups.

The added section section_name applies only to documents indexed after this operation. Thus for the change to take effect, you must manually re-index any existing documents that contain the tag.

The index is *not* rebuilt by this statement.

Note:  This ALTER INDEX operation applies only to CONTEXT CTXRULE indexes. It does not apply to CTXCAT indexes.

See Also:  "Notes" on page 1-15.

ADD STOP SECTION tag
Dynamically adds the stop section identified by tag to the existing index. As stop sections apply only to automatic sectioning of XML documents, the index must use the AUTO SECTION GROUP section group. The tag you specify must be case sensitive and unique within the automatic section group or else ALTER INDEX raises an error.

The added stop section tag applies only to documents indexed after this operation. For the change to affect previously indexed documents, you must explicitly re-index the documents that contain the tag.

The text within a stop section is always searchable.

The number of stop sections you can add is unlimited.

The index is *not* rebuilt by this statement.

See Also:  "Notes" on page 1-15

Note:  This ALTER INDEX operation applies only to CONTEXT indexes. It does not apply to CTXCAT indexes.

PARALLEL n
Optionally specify with n the parallel degree for parallel indexing. This parameter is supported only when you use SYNC, REPLACE, and RESUME in paramstring. The actual degree of parallelism might be smaller depending on your resources.

Parallel indexing can speed up indexing when you have large amounts of data to index and when your operating system supports multiple CPUs.
ALTER INDEX Examples

**Resuming Failed Index**
The following statement resumes the indexing operation on `newsindex` with 2 megabytes of memory:

```
ALTER INDEX newsindex REBUILD PARAMETERS('resume memory 2M');
```

**Rebuilding an Index**
The following statement rebuilds the index, replacing the stoplist preference with `new_stop`.

```
ALTER INDEX newsindex REBUILD PARAMETERS('replace stoplist new_stop');
```

**Rebuilding a Partitioned Index**
The following example creates a partitioned text table, populates it, and creates a partitioned index. It then adds a new partition to the table and then rebuilds the index with `ALTER INDEX` as follows:

```
PROMPT create partitioned table and populate it

create table part_tab (a int, b varchar2(40)) partition by range(a)
(partition p_tab1 values less than (10),
 partition p_tab2 values less than (20),
 partition p_tab3 values less than (30));

insert into part_tab values (1, 'Actinidia deliciosa');
insert into part_tab values (8, 'Distictis buccinatoria');
insert into part_tab values (12, 'Actinidia quinata');
insert into part_tab values (18, 'Distictis Rivers');
insert into part_tab values (21, 'pandorea jasminoides Lady Di');
insert into part_tab values (28, 'pandorea rosea');

commit;

PROMPT create partitioned index
create index part_idx on part_tab(b) indextype is ctxsys.context
local (partition p_idx1, partition p_idx2, partition p_idx3);

PROMPT add a partition and populate it
alter table part_tab add partition p_tab4 values less than (40);
insert into part_tab values (32, 'passiflora citrina');
insert into part_tab values (33, 'passiflora alatocaerulea');

commit;
```

The following statement rebuilds the index in the newly populated partition. In general, the index partition name for a newly added partition is the same as the table partition name, unless the name has already been used. In this case, Oracle Text generates a new name.

```
alter index part_idx rebuild partition p_tab4;
```

The following statement queries the table for the two hits in the newly added partition:

```
select * from part_tab where contains(b, 'passiflora') >0;
```

The following statement queries the newly added partition directly:
select * from part_tab partition (p_tab4) where contains(b,'passiflora') >0;

**Replacing Index Metadata: Changing Single-Lexer to Multilexer**

The following example demonstrates how an application can migrate from single-language documents (English) to multi-language documents (English and Spanish) by replacing the index metadata for the lexer.

REM create a simple table, which stores only english (American) text

create table simple (text varchar2(80));
insert into simple values ('the quick brown fox');
commit;

REM we’ll create a simple lexer to lex this english text

begin
  ctx_ddl.create_preference('us_lexer','basic_lexer');
end;
/

REM create a text index on the simple table
create index simple_idx on simple(text)
indextype is ctxsys.context parameters ('lexer us_lexer');

REM we can query easily
select * from simple where contains(text, 'fox')>0;

REM now suppose we want to start accepting spanish documents.
REM first we have to extend the table with a language column
alter table simple add (lang varchar2(10) default 'us');

REM now let’s create a spanish lexer,
begin
  ctx_ddl.create_preference('e_lexer','basic_lexer');
  ctx_ddl.set_attribute('e_lexer','base_letter','yes');
end;
/

REM Then we create a multi-lexer incorporating our english and spanish lexers.
REM Note that the DEFAULT lexer is the exact same lexer that we have already
REM indexed all the documents with.
begin
  ctx_ddl.create_preference('m_lexer','multi_lexer');
  ctx_ddl.add_sub_lexer('m_lexer','default','us_lexer');
  ctx_ddl.add_sub_lexer('m_lexer','spanish','e_lexer');
end;
/

REM now let’s replace our metadata
alter index simple_idx rebuild
parameters ('replace metadata language column lang lexer m_lexer');

REM we’re ready for some spanish data. Note that we could have inserted
REM this BEFORE the alter index, as long as we didn’t SYNC.
insert into simple values ('el zorro marron rapido', 'e');
commit;
exec ctx_ddl.sync_index('simple_idx');
REM now we can query the spanish data with base lettering:
select * from simple where contains(text, 'rapido')>0;
Optimizing the Index
To optimize your index, use CTX_DDL.OPTIMIZE_INDEX.

Synchronizing the Index
To synchronize your index, use CTX_DDL.SYNC_INDEX.

Adding a Zone Section
To add to the index the zone section author identified by the tag <author>, enter the following statement:

```
ALTER INDEX myindex REBUILD PARAMETERS('add zone section author tag author');
```

Adding a Stop Section
To add a stop section identified by tag <fluff> to the index that uses the AUTO_SECTION_GROUP, enter the following statement:

```
ALTER INDEX myindex REBUILD PARAMETERS('add stop section fluff');
```

Adding an Attribute Section
Assume that the following text appears in an XML document:

```
<book title="Tale of Two Cities">It was the best of times.</book>
```

and that you want to create a separate section for the title attribute and you want to name the new attribute section booktitle. To do so, enter the following statement:

```
ALTER INDEX myindex REBUILD PARAMETERS('add attr section booktitle tag title@book');
```

Using Flashback Queries
If a Text query is flashed back to a point before an ALTER INDEX statement was issued on the Text index for which the query is being run, then:

- The query optimizer will not choose the index access path for that given index because the index is treated according to its creation time with ALTER INDEX. Therefore, to the query optimizer, the index is perceived not to exist.
- The functional processing of the Text operator will fail with ORA-01466 or ORA-08176 errors if the ALTER INDEX statement involves re-creation of DR$ index tables.

To work around this issue, use the DBMS_FLASHBACK package. For example:

```
EXEC dbms_flashback.enable_at_system_change_number(:scn);
SELECT id from documents WHERE CONTAINS(text, 'oracle')>0;
EXEC dbms_flashback.disable;
```

Note: In previous releases, flashback Text queries using AS OF predicates with Text operators such as CONTAINS and CATSEARCH are not supported.

See Also: "Using DBMS_FLASHBACK Package" in Oracle Database Advanced Application Developer's Guide
Notes

Add Section Constraints
Before altering the index section information, Oracle Text checks the new section against the existing sections to ensure that all validity constraints are met. These constraints are the same for adding a section to a section group with the CTX_DDL PL/SQL package and are as follows:

- You cannot add zone, field, or stop sections to a NULL SECTION GROUP.
- You cannot add zone, field, or attribute sections to an automatic section group.
- You cannot add attribute sections to anything other than XML section groups.
- You cannot have the same tag for two different sections.
- Section names for zone, field, and attribute sections cannot intersect.
- You cannot exceed 64 fields per section.
- You cannot add stop sections to basic, HTML, XML, or news section groups.
- SENTENCE and PARAGRAPH are reserved section names.

Related Topics

CTX_DDL.SYNC_INDEX in Chapter 7, "CTX_DDL Package"
CTX_DDL.OPTIMIZE_INDEX in Chapter 7, "CTX_DDL Package"
CREATE INDEX on page 1-36
ALTER TABLE: Supported Partitioning Statements

**Purpose**

Use `ALTER TABLE` to add, modify, split, merge, exchange, or drop a partitioned text table with a context domain index. The following sections describe some of the `ALTER TABLE` operations.

**Modify Partition Syntax**

**Unusable Local Indexes**

```
ALTER TABLE [schema.]table MODIFY PARTITION partition UNUSABLE LOCAL INDEXES
```

Marks the index partition corresponding to the given table partition UNUSABLE. You might mark an index partition unusable before you rebuild the index partition as described in "Rebuild Unusable Local Indexes".

If the index partition is not marked unusable, then the rebuild command returns without actually rebuilding the local index partition.

**Rebuild Unusable Local Indexes**

```
ALTER TABLE [schema.]table MODIFY PARTITION partition REBUILD UNUSABLE LOCAL INDEXES
```

Rebuilds the index partition corresponding to the specified table partition that has an UNUSABLE status.

**Note:** If the index partition status is already VALID before you enter this command, then this command does not rebuild the index partition. Do not depend on this command to rebuild the index partition unless the index partition status is UNUSABLE.

**Add Partition Syntax**

```
ALTER TABLE [schema.]table ADD PARTITION [partition]
VALUES LESS THAN {value_list} [partition_description]
```

Adds a new partition to the high end of a range partitioned table.

To add a partition to the beginning or to the middle of the table, use `ALTER TABLE SPLIT PARTITION`.

The newly added table partition is always empty, and the context domain index (if any) status for this partition is always VALID. After doing DML, if you want to synchronize or optimize this newly added index partition, then you must look up the index partition name and enter the `ALTER INDEX REBUILD PARTITION` statement.
For this newly added partition, index partition name is usually the same as the table partition name, but if the table partition name is already used by another index partition, the system assigns a name in the form of SYS_Pn.

By querying the USER_IND_PARTITIONS view and comparing the HIGH_VALUE field, you can determine the index partition name for the newly added partition.

**Merge Partition Syntax**

```
ALTER TABLE [schema.]table
MERGE PARTITIONS partition1, partition2
[INTO PARTITION [new_partition] [partition_description]]
[UPDATE GLOBAL INDEXES]
```

Applies only to a range partition. This command merges the contents of two adjacent partitions into a new partition and then drops the original two partitions. If the resulting partition is non-empty, then the corresponding local domain index partition is marked UNUSABLE. You can use ALTER TABLE MODIFY PARTITION to rebuild the partition index.

For a global, nonpartitioned index, if you perform the merge operation without an UPDATE GLOBAL INDEXES clause, then the resulting index (if not NULL) will be invalid and must be rebuilt. If you specify the UPDATE GLOBAL INDEXES clause after the operation and the sync type is manual, then the index will be valid, but you still must synchronize the index with CTX_DDL.SYNC_INDEX for the update to take place.

The naming convention for the resulting index partition is the same as in ALTER TABLE ADD PARTITION.

**Split Partition Syntax**

```
ALTER TABLE [schema.]table
SPLIT PARTITION partition_name_old
AT (value_list)
[into (partition_description, partition_description)]
[parallel_clause]
[UPDATE GLOBAL INDEXES]
```

Applies only to range partition. This command divides a table partition into two partitions, thus adding a new partition to the table. The local corresponding index partitions will be marked UNUSABLE if the corresponding table partitions are non-empty. Use ALTER TABLE MODIFY PARTITION to rebuild the partition indexes.

For a global, nonpartitioned index, if you perform the split operation without an UPDATE GLOBAL INDEXES clause, then the resulting index (if not NULL) will be invalid and must be rebuilt. If you specify the UPDATE GLOBAL INDEXES clause after the operation and the sync type is manual, then the index will be valid, but you still must synchronize the index with CTX_DDL.SYNC_INDEX for the update to take place.

The naming convention for the two resulting index partition is the same as in ALTER TABLE ADD PARTITION.

**Exchange Partition Syntax**

```
ALTER TABLE [schema.]table EXCHANGE PARTITION partition WITH TABLE table
[INCLUDING|EXCLUDING INDEXES]
[WITH|WITHOUT VALIDATION]
[EXCEPTIONS INTO [schema.]table]
```
[UPDATE GLOBAL INDEXES]

Converts a partition to a nonpartitioned table, and converts a table to a partition of a partitioned table by exchanging their data segments. Rowids are preserved.

If EXCLUDING INDEXES is specified, all the context indexes corresponding to the partition and all the indexes on the exchanged table are marked as UNUSABLE. To rebuild the new index partition in this case, enter ALTER TABLE MODIFY PARTITION.

If INCLUDING INDEXES is specified, then for every local domain index on the partitioned table, there must be a nonpartitioned domain index on the nonpartitioned table. The local index partitions are exchanged with the corresponding regular indexes.

For a global, nonpartitioned index, if you perform the exchange operation without an UPDATE GLOBAL INDEXES clause, then the resulting index (if not NULL) will be invalid and must be rebuilt. If you specify the UPDATE GLOBAL INDEXES clause after the operation and the sync type is manual, then the index will be valid, but you still must synchronize the index with CTX_DDL.SYNC_INDEX for the update to take place.

Field Sections
Field section queries might not work the same if the nonpartitioned index and local index use different section IDs for the same field section.

Storage
Storage is not changed. So if the index on the nonpartitioned table $I table was in tablespace XYZ, then after the exchange partition, it will still be in tablespace XYZ, but now it is the $I table for an index partition.

Storage preferences are not switched, so if you switch and then rebuild the index, then the table may be created in a different location.

Restrictions
Both indexes must be equivalent. They must use the same objects and the same settings for each object. Note that Oracle Text only checks that the indexes are using the same object. But they should use the same exact everything.

No index object can be partitioned, that is, when the user has used the storage object to partition the $I, $N tables.

If either index or index partition does not meet all these restrictions an error is raised and both the index and index partition will be INVALID. The user needs to manually rebuild both index and index partition using ALTER INDEX REBUILD.

Truncate Partition Syntax

ALTER TABLE [schema.]table TRUNCATE PARTITION [DROP|REUSE STORAGE] [UPDATE GLOBAL INDEXES]

Removes all rows from a partition in a table. Corresponding CONTEXT index partitions are also removed.

For a global, nonpartitioned index, if you perform the truncate operation without an UPDATE GLOBAL INDEXES clause, then the resulting index (if not NULL) will be invalid and must be rebuilt. If you specify the UPDATE GLOBAL INDEXES clause after the operation, the index will be valid.
**ALTER TABLE Examples**

### Global Index on Partitioned Table Examples

The following example creates a range-partitioned table with three partitions. Each partition is populated with two rows. A global, nonpartitioned CONTEXT index is then created. To demonstrate the UPDATE GLOBAL INDEXES clause, the partitions are split and merged with an index synchronization.

```sql
create table tdrexglb_part(a int, b varchar2(40)) partition by range(a)
(partition p1 values less than (10),
 partition p2 values less than (20),
 partition p3 values less than (30));

insert into tdrexglb_part values (1,'row1');
insert into tdrexglb_part values (8,'row2');
insert into tdrexglb_part values (11,'row11');
insert into tdrexglb_part values (18,'row18');
insert into tdrexglb_part values (21,'row21');
insert into tdrexglb_part values (28,'row28');
commit;
create index tdrexglb_parti on tdrexglb_part(b) indextype is ctxsys.context;
create table tdrexglb(a int, b varchar2(40));
insert into tdrexglb values(20,'newrow20');
commit;
PROMPT make sure query works
select * from tdrexglb_part where contains(b,'row18') >0;
PROMPT split partition
alter table tdrexglb_part split partition p2 at (15) into
(partition p21, partition p22) update global indexes;
PROMPT before sync
select * from tdrexglb_part where contains(b,'row11') >0;
select * from tdrexglb_part where contains(b,'row18') >0;
exec ctx_ddl.sync_index('tdrexglb_parti')
PROMPT after sync
select * from tdrexglb_part where contains(b,'row11') >0;
select * from tdrexglb_part where contains(b,'row18') >0;
PROMPT merge partition
alter table tdrexglb_part merge partitions p22, p3
into partition pnew3 update global indexes;
PROMPT before sync
select * from tdrexglb_part where contains(b,'row18') >0;
select * from tdrexglb_part where contains(b,'row28') >0;
exec ctx_ddl.sync_index('tdrexglb_parti');
PROMPT after sync
select * from tdrexglb_part where contains(b,'row18') >0;
select * from tdrexglb_part where contains(b,'row28') >0;
```

---

**Oracle Text SQL Statements and Operators** 1-19
PROMPT drop partition
alter table tdrexglb_part drop partition p1 update global indexes;

PROMPT before sync
select * from tdrexglb_part where contains(b,'row1') >0;
exec ctx_ddl.sync_index('tdrexglb_parti');

PROMPT after sync
select * from tdrexglb_part where contains(b,'row1') >0;

PROMPT exchange partition
alter table tdrexglb_part exchange partition pnew3 with table tdrexglb update global indexes;

PROMPT before sync
select * from tdrexglb_part where contains(b,'newrow20') >0;
select * from tdrexglb_part where contains(b,'row28') >0;
exec ctx_ddl.sync_index('tdrexglb_parti');

PROMPT after sync
select * from tdrexglb_part where contains(b,'newrow20') >0;
select * from tdrexglb_part where contains(b,'row28') >0;

PROMPT move table partition
alter table tdrexglb_part move partition p21 update global indexes;

PROMPT before sync
select * from tdrexglb_part where contains(b,'row11') >0;
exec ctx_ddl.sync_index('tdrexglb_parti');

PROMPT after sync
select * from tdrexglb_part where contains(b,'row11') >0;

PROMPT truncate table partition
alter table tdrexglb_part truncate partition p21 update global indexes;
update global indexes;
CATSEARCH

Use the CATSEARCH operator to search CTXCAT indexes. Use this operator in the WHERE clause of a SELECT statement.

The CATSEARCH operator also supports database links. You can identify a remote table or materialized view by appending @dblink to the end of its name. The dblink must be a complete or partial name for a database link to the database containing the remote table or materialized view. (Indexing of remote views is not supported.)

The grammar of this operator is called CTXCAT. You can also use the CONTEXT grammar if your search criteria require special functionality, such as thesaurus, fuzzy matching, proximity searching or stemming. To utilize the CONTEXT grammar, use the Query Template Specification in the text_query parameter as described in this section.

About Performance

Use the CATSEARCH operator with a CTXCAT index mainly to improve mixed query performance. Specify your text query condition with text_query and your structured condition with structured_query.

Internally, Oracle Text uses a combined b-tree index on text and structured columns to quickly produce results satisfying the query.

Limitations

If the optimizer chooses to use the functional query invocation, then your query will fail. The optimizer might choose functional invocation when your structured clause is highly selective.

The structured_query argument of the CATSEARCH operator must reference columns used during CREATE INDEX sets; otherwise, error DRG-10845 will be raised. For example, the error will be raised if you issue a CATSEARCH query on a view created on top of a table with CTXCAT index on it, and the name of the logical column on the view is different from the actual column name on the physical table. The columns referenced by the structured_query argument of CATSEARCH operator must be the physical column name used during CREATE INDEX sets, not the logical column on the view.

Syntax

```
CATSEARCH(
    [schema.]column,
    text_query   [VARCHAR2|CLOB],
    structured_query VARCHAR2,
    RETURN NUMBER;

[schema.]column
Specify the text column to be searched on. This column must have a CTXCAT index associated with it.

text_query
Specify one of the following to define your search in column.

-  CATSEARCH query operations
-  Query Template Specification (for using CONTEXT grammar)
```
**CATSEARCH query operations**

The **CATSEARCH** operator supports only the following query operations:

- Logical **AND**
- Logical **OR** (|)
- Logical **NOT** (-)
- " " (quoted phrases)
- Wildcarding

The following table provides the syntax for these operators.

<table>
<thead>
<tr>
<th>Table 1–2</th>
<th>CATSEARCH Query Operators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Operation</strong></td>
<td><strong>Syntax</strong></td>
</tr>
<tr>
<td>Logical <strong>AND</strong></td>
<td>a b c</td>
</tr>
<tr>
<td>Logical <strong>OR</strong></td>
<td>a</td>
</tr>
<tr>
<td>Logical <strong>NOT</strong></td>
<td>a - b</td>
</tr>
<tr>
<td>Hyphen with no space</td>
<td>a-b</td>
</tr>
<tr>
<td>&quot; &quot; (quoted phrases)</td>
<td>&quot;a b c&quot;</td>
</tr>
<tr>
<td>( ) (Parentheses group operations)</td>
<td>(A B)</td>
</tr>
<tr>
<td>Wildcard (right and double truncated)</td>
<td>term*</td>
</tr>
</tbody>
</table>

The following limitations apply to these operators:

- The left-hand side (the column name) must be a column named in at least one of the indexes of the index set.
- The left-hand side must be a plain column name. Functions and expressions are not allowed.
- The right-hand side must be composed of literal values. Functions, expressions, other columns, and subselects are not allowed.
Multiple criteria can be combined with **AND**. Note that **OR** is not supported.

When querying a remote table through a database link, the database link must be specified for **CATSEARCH** as well as for the table being queried.

For example, these expressions are supported:

```sql
catsearch(text, 'dog', 'foo > 15')
catsearch(text, 'dog', 'bar = ''SMITH''')
catsearch(text, 'dog', 'foo between 1 and 15')
catsearch(text, 'dog', 'foo = 1 and abc = 123')
catsearch@remote(text, 'dog', 'foo = 1 and abc = 123')
```

And these expressions are not supported:

```sql
catsearch(text, 'dog', 'upper(bar) = ''A''')
catsearch(text, 'dog', 'bar LIKE ''A%''')
catsearch(text, 'dog', 'foo = abc')
catsearch(text, 'dog', 'foo = 1 or abc = 3')
```

**Query Template Specification**

Specify a marked-up string that specifies a query template. Specify one of the following templates:

- Query rewrite, used to expand a query string into different versions
- Progressive relaxation, used to progressively enter less restrictive versions of a query to increase recall
- Alternate grammar, used to specify **CONTAINS** operators (See **CONTEXT Query Grammar Examples**)
- Alternate language, used to specify alternate query language
- Alternate scoring, used to specify alternate scoring algorithms

**See Also:** The **text_query** parameter description for **CONTAINS** on page 1-28 for more information about the syntax for these query templates.

**structured_query**

Specify the structured conditions and the **ORDER BY** clause. There must exist an index for any column you specify. For example, if you specify `'category_id=1 order by bid_close'`, you must have an index for `'category_id, bid_close'` as specified with **CTX_DDL.ADD_INDEX**.

With **structured_query**, you can use standard SQL syntax with only the following operators:

- `=`
- `<=`
- `>=`
- `>`
- `<`
- `IN`
- `BETWEEN`
- `AND` (to combine two or more clauses)
Examples

1. Create the Table

The following statement creates the table to be indexed.

```sql
CREATE TABLE auction (category_id number primary key, title varchar2(20), bid_close date);
```

The following table inserts the values into the table:

```sql
INSERT INTO auction values(1, 'Sony CD Player', '20-FEB-2000');
INSERT INTO auction values(2, 'Sony CD Player', '24-FEB-2000');
INSERT INTO auction values(3, 'Pioneer DVD Player', '25-FEB-2000');
INSERT INTO auction values(4, 'Sony CD Player', '25-FEB-2000');
INSERT INTO auction values(5, 'Bose Speaker', '22-FEB-2000');
INSERT INTO auction values(6, 'Tascam CD Burner', '25-FEB-2000');
INSERT INTO auction values(7, 'Nikon digital camera', '22-FEB-2000');
INSERT INTO auction values(8, 'Canon digital camera', '26-FEB-2000');
```

1. Create the CTXCAT Index

The following statements create the CTXCAT index:

```sql
begin
  ctx_ddl.create_index_set('auction_iset');
  ctx_ddl.add_index('auction_iset','bid_close');
end;
/
CREATE INDEX auction_titlex ON auction(title) INDEXTYPE IS CTXSYS.CTXCAT
PARAMETERS ('index set auction_iset');
```

1. Query the Table

A typical query with CATSEARCH might include a structured clause as follows to find all rows that contain the word `camera` ordered by `bid_close`:

```sql
SELECT * FROM auction WHERE CATSEARCH(title, 'camera', 'order by bid_close desc')> 0;
```

<table>
<thead>
<tr>
<th>CATEGORY_ID</th>
<th>TITLE</th>
<th>BID_CLOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Canon digital camera</td>
<td>26-FEB-00</td>
</tr>
<tr>
<td>7</td>
<td>Nikon digital camera</td>
<td>22-FEB-00</td>
</tr>
</tbody>
</table>

The following query finds all rows that contain the phrase `Sony CD Player` and that have a bid close date of February 20, 2000:

```sql
SELECT * FROM auction WHERE CATSEARCH(title, '"Sony CD Player"', 'bid_close=''20-FEB-00''')> 0;
```

<table>
<thead>
<tr>
<th>CATEGORY_ID</th>
<th>TITLE</th>
<th>BID_CLOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sony CD Player</td>
<td>20-FEB-00</td>
</tr>
</tbody>
</table>

The following query finds all rows with the terms `Sony` and `CD` and `Player`:

```sql
SELECT * FROM auction WHERE CATSEARCH(title, 'Sony CD Player', 'order by bid_close desc')> 0;
```
The following query finds all rows with the term CD and not Player:

```
SELECT * FROM auction WHERE CATSEARCH(title, 'CD - Player', 'order by bid_close desc') > 0;
```

<table>
<thead>
<tr>
<th>CATEGORY_ID</th>
<th>TITLE</th>
<th>BID_CLOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Sony CD Player</td>
<td>25-FEB-00</td>
</tr>
<tr>
<td>2</td>
<td>Sony CD Player</td>
<td>24-FEB-00</td>
</tr>
<tr>
<td>1</td>
<td>Sony CD Player</td>
<td>20-FEB-00</td>
</tr>
</tbody>
</table>

The following query finds all rows with the terms CD or DVD or Speaker:

```
SELECT * FROM auction WHERE CATSEARCH(title, 'CD | DVD | Speaker', 'order by bid_close desc') > 0;
```

<table>
<thead>
<tr>
<th>CATEGORY_ID</th>
<th>TITLE</th>
<th>BID_CLOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Tascam CD Burner</td>
<td>25-FEB-00</td>
</tr>
<tr>
<td>3</td>
<td>Pioneer DVD Player</td>
<td>25-FEB-00</td>
</tr>
<tr>
<td>4</td>
<td>Sony CD Player</td>
<td>25-FEB-00</td>
</tr>
<tr>
<td>6</td>
<td>Tascam CD Burner</td>
<td>25-FEB-00</td>
</tr>
<tr>
<td>2</td>
<td>Sony CD Player</td>
<td>24-FEB-00</td>
</tr>
<tr>
<td>5</td>
<td>Bose Speaker</td>
<td>22-FEB-00</td>
</tr>
<tr>
<td>1</td>
<td>Sony CD Player</td>
<td>20-FEB-00</td>
</tr>
</tbody>
</table>

The following query finds all rows that are about audio equipment:

```
SELECT * FROM auction WHERE CATSEARCH(title, 'ABOUT(audio equipment)', NULL) > 0;
```

**CONTEXT Query Grammar Examples**

The following examples show how to specify the CONTEXT grammar in CATSEARCH queries using the template feature.

**PROMPT**

**PROMPT fuzzy: query = ?test**

**PROMPT should match all fuzzy variations of test (for example, text)**

```
select pk||' ==> '||text from test
where catsearch|text,
'<query>
 <textquery grammar="context">
 ?test
 </textquery>
</query>'','')>0
order by pk;
```

**PROMPT**

**PROMPT fuzzy: query = !sail**

**PROMPT should match all soundex variations of bot (for example, sell)**

```
select pk||' ==> '||text from test
where catsearch|text,
'<query>
 <textquery grammar="context">
 !sail
 </textquery>
</query>'','')>0
order by pk;
```
The following example shows a field section search against a CTXCAT index using CONTEXT grammar by means of a query template in a CATSEARCH query.

```sql
-- Create and populate table
create table BOOKS (ID number, INFO varchar2(200), PUBDATE DATE);
insert into BOOKS values(1, '<author>NOAM CHOMSKY</author><subject>CIVIL RIGHTS</subject><language>ENGLISH</language><publisher>MIT PRESS</publisher>', '01-NOV-2003');
insert into BOOKS values(2, '<author>NICANOR PARRA</author><subject>POEMS AND ANTIPOEMS</subject><language>SPANISH</language><publisher>VASQUEZ</publisher>', '01-JAN-2001');
insert into BOOKS values(1, '<author>LUC SANTE</author><subject>XML DATABASE</subject><language>FRENCH</language><publisher>FREE PRESS</publisher>', '15-MAY-2002');
commit;

-- Create index set and section group
exec ctx_ddl.create_index_set('BOOK_INDEX_SET');
exec ctx_ddl.add_index('BOOKSET','PUBDATE');
exec ctx_ddl.create_section_group('BOOK_SECTION_GROUP', 'BASIC_SECTION_GROUP');
exec ctx_ddl.add_field_section('BOOK_SECTION_GROUP','AUTHOR','AUTHOR');
exec ctx_ddl.add_field_section('BOOK_SECTION_GROUP','SUBJECT','SUBJECT');
exec ctx_ddl.add_field_section('BOOK_SECTION_GROUP','LANGUAGE','LANGUAGE');
exec ctx_ddl.add_field_section('BOOK_SECTION_GROUP','PUBLISHER','PUBLISHER');

-- Create index
create index books_index on books(info) indextype is ctxsys.ctxcat
parameters('index set book_index_set section group book_section_group');

-- Use the index
-- Note that: even though CTXCAT index can be created with field sections, it cannot be accessed using CTXCAT grammar (default for CATSEARCH).
-- We need to use query template with CONTEXT grammar to access field sections with CATSEARCH
select id, info from books
where catsearch(info,
'NOAM within author and english within language
', '');
Related Topics

Syntax for CTXCAT Index Type in this chapter.

Oracle Text Application Developer’s Guide
CONTAINS

Use the CONTAINS operator in the WHERE clause of a SELECT statement to specify the query expression for a Text query.

The CONTAINS operator also supports database links. You can identify a remote table or materialized view by appending @dblink to the end of its name. The dblink must be a complete or partial name for a database link to the database containing the remote table or materialized view. (Querying of remote views is not supported.)

CONTAINS returns a relevance score for every row selected. Obtain this score with the SCORE operator.

The grammar for this operator is called the CONTEXT grammar. You can also use CTXCAT grammar if your application works better with simpler syntax. To do so, use the Query Template Specification in the text_query parameter as described in this section.

**See Also:** "The CONTEXT Grammar" topic in Oracle Text Application Developer’s Guide

**Syntax**

```
CONTAINS(
    [schema.]column,
    text_query [VARCHAR2|CLOB]
    [,label NUMBER])
RETURN NUMBER;
```

**[schema.]column**
Specify the text column to be searched on. This column must have a Text index associated with it.

**text_query**
Specify one of the following:

- The query expression that defines your search in column.
- A marked-up document that specifies a query template. Use one of the following templates:

**Query Rewrite Template**

Use this template to automatically write different versions of a query before you submit the query to Oracle Text. This is useful when you need to maximize the recall of a user query. For example, you can program your application to expand a single phrase query of 'cat dog' into the following queries:

```
{cat} {dog}
{cat} ; {dog}
{cat} AND {dog}
{cat} ACCUM {dog}
```

These queries are submitted as one query and results are returned with no duplication. In this example, the query returns documents that contain the phrase *cat dog* as well as documents in which *cat* is near *dog*, and documents that have *cat* and *dog*.

This is done with the following template:
The operator *TRANSFORM* is used to specify the rewrite rules and has the following syntax (note that it uses double parentheses):

```
TRANSFORM((terms, prefix, suffix, connector))
```

### Table 1–3 TRANSFORM Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term</td>
<td>Specify the type of terms to be produced from the original query. Specify either \textit{TOKENS} or \textit{THEMES}. Specifying \textit{THEMES} requires an installed knowledge base. A knowledge base may or may not have been installed with Oracle Text. For more information, see the Oracle Text Application Developer's Guide.</td>
</tr>
<tr>
<td>prefix</td>
<td>Specify the literal string to be prepended to all terms.</td>
</tr>
<tr>
<td>suffix</td>
<td>Specify the literal string to be appended to all terms.</td>
</tr>
<tr>
<td>connector</td>
<td>Specify the literal string to connect all terms after applying prefix and suffix.</td>
</tr>
</tbody>
</table>

**Note:** An error will be raised if the input Text query string specified in the Query Rewrite Template with \textit{TRANSFORM} rules contains any Oracle Text query operators (such as \textit{AND}, \textit{OR}, or \textit{SOUNDEX}). Also, any special characters (such as \% or \$) in the input Text query string must be escaped, or an error is raised.

### Query Relaxation Template

Use this template to progressively relax your query. Progressive relaxation is when you increase recall by progressively issuing less restrictive versions of a query, so that your application can return an appropriate number of hits to the user.

For example, the query of \textit{black pen} can be progressively relaxed to:

- black pen
- black NEAR pen
- black AND pen
- black ACCUM pen

This is done with the following template

```
<query>
    <textquery lang="ENGLISH" grammar="CONTEXT"> cat dog
    <progression>
        <seq><rewrite>transform((TOKENS, "{", ",", " "]))</rewrite></seq>
        <seq><rewrite>transform((TOKENS, "{", ",", "; ")</rewrite></seq>
        <seq><rewrite>transform((TOKENS, "{", ",", "AND")</rewrite></seq>
        <seq><rewrite>transform((TOKENS, "{", ",", "ACCUM")</rewrite></seq>
    </progression>
</textquery>
```

```
<score datatype="INTEGER" algorithm="COUNT"/>
```

```
</query>
```
Alternate Grammar Template

Use this template to specify an alternate grammar, such as CONTEXT or CATSEARCH. Specifying an alternate grammar enables you to enter queries using different syntax and operators.

For example, with CATSEARCH, enter ABOUT queries using the CONTEXT grammar. Likewise with CONTAINS, enter logical queries using the simplified CATSEARCH syntax.

The phrase ‘dog cat mouse’ is interpreted as a phrase in CONTAINS. However, with CATSEARCH this is equivalent to a AND query of ‘dog AND cat AND mouse’. To specify that CONTAINS use the alternate grammar, enter the following template:

```xml
<query>
  <textquery grammar="CTXCAT">dog cat mouse</textquery>
  <score datatype="integer"/>
</query>
```

Alternate Language Template

Use this template to specify an alternate language.

```xml
<query><textquery lang="french">bon soir</textquery></query>
```

Alternative Scoring Template

Use this template to specify an alternative scoring algorithm.

The following example specifies that the query use the CONTEXT grammar and return integer scores using the COUNT algorithm. This algorithm returns score as the number of query occurrences in the document.

```xml
<query>
  <textquery grammar="CONTEXT" lang="english"> mustang </textquery>
  <score datatype="INTEGER" algorithm="COUNT"/>
</query>
```

The following example uses the normalization_expr attribute to add SDATA(price) into the score returned by the query, and uses it as the final score.

```xml
<query>
  <textquery grammar="CONTEXT" lang="english"> DEFINESCORE(dog, RELEVANCE) and  cat </textquery>
  <score  algorithm="COUNT"  normalization_expr ="doc_score+ SDATA(price)"/>
</query>
```

The normalization_expr attribute is used only with the alternate scoring template, and is an arithmetic expression that consists of:

- Arithmetic operators: + - * / . The operator precedence is the same as that for SQL operator precedence.
- Grouping operators: (). Parentheses can be used to alter the precedence of the arithmetic operators.
■ Absolute function: ABS(n) returns the absolute value of n; where n is any expression that returns a number.

■ Logarithmic function: LOG(n): returns the base-10 logarithmic value of n; where n is any expression that returns a number.

■ Predefined components: The doc_score predefined component can be used to return the initial query score of a particular document.

■ SDATA component: SDATA(name) returns the value of the SDATA with the specified name as the score.
  – Only SDATA with NUMBER or DATE datatype is allowed. An error is raised otherwise.
  – The data string and the SDATA name are case-insensitive.
  – As an SDATA section value can be NULL, any expression with NULL SDATA section value is evaluated as 0. For example: the normalizationExpr "doc_score + SDATA(price)" will be evaluated to 0 if SDATA(price) for a given document has NULL value.

■ Numeric literals: Any number literal that conforms to SQL's pattern of NUMBER literal and is within the range of the double precision floating point (-3.4e38 to 3.4e38).

■ Date Literals: Date literals must be enclosed with DATE(). Only the following format is allowed: YYYY-MM-DD or YYYY-MM-DD HH24:MI:SS. For example: DATE(2005-11-08). Consistent with SQL, if no time is specified, then 00:00:00 is assumed.

The normalizationExpr attribute overrides the algorithm attribute. That is, if algorithm is set to COUNT, and the user also specifies normalizationExpr, then the score will not be count, but the calculated score based on the normalizationExpr.

If the score (either from algorithm = COUNT or normalizationExpr = ...) is internally calculated to be greater than 100, then it will be set to 100.

If the query relaxation template is used, the score will be further normalized in such a way that documents returned from higher sequences will always have higher scores than documents returned from sequence(s) below.

**DATE Literal Restrictions**

Only the minus (-) operator is allowed between date-type data (DATE literals and date-type SDATA). Using other operators will result in an error. Subtracting two date-type data will produce a number (float) that represents the difference in number of days between the two dates. For example, the following expression is allowed:

\[ \text{SDATA(dob)} - \text{DATE(2005-11-08)} \]

and the following expression is not allowed:

\[ \text{SDATA(dob)} + \text{DATE(2005-11-08)} \]

The plus (+) and minus (-) operators are allowed between numeric data and date type of data. The number operand is interpreted as the number or fraction of day(s). For example, the following expression is allowed:

\[ \text{DATE(2005-11-08)} + 1 = 9 \text{ NOV 2005} \]

and the following expression is not allowed:

\[ \text{DATE(2005-11-08)} \times 3 \text{ = ERROR} \]
**Template Attribute Values**

Table 1–4 gives the possible values for template attributes:

<table>
<thead>
<tr>
<th>Tag Attribute</th>
<th>Description</th>
<th>Possible Values</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>grammar=</td>
<td>Specify the grammar of the query.</td>
<td>CONTEXT, CTXCAT</td>
<td>Returns score as integer between 0 and 100.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Returns score as its high precision floating point number between 0 and 100.</td>
</tr>
<tr>
<td>datatype=</td>
<td>Specify the type of number returned as score.</td>
<td>INTEGER, FLOAT</td>
<td></td>
</tr>
<tr>
<td>algorithm=</td>
<td>Specify the scoring algorithm to use.</td>
<td>DEFAULT, COUNT</td>
<td>Returns scores as the number of occurrences in document.</td>
</tr>
<tr>
<td>lang=</td>
<td>Specify the language name.</td>
<td>Any language supported by Oracle Database. See Oracle Database Globalization Support Guide.</td>
<td></td>
</tr>
</tbody>
</table>

**Template Grammar Definition**

The query template interface is an XML document. Its grammar is defined with the following XML DTD:

```xml
<!ELEMENT query (textquery, score?)>
<!ELEMENT textquery (#PCDATA|progression)*>  
<!ELEMENT progression (seq)+>
<!ELEMENT seq (#PCDATA|rewrite)>  
<!ELEMENT rewrite (#PCDATA)>  
<!ELEMENT score EMPTY>  
<!ATTLIST textquery grammar (context | ctxcat) #IMPLIED>  
<!ATTLIST textquery language CDATA #IMPLIED>  
<!ATTLIST score datatype (integer | float) "integer">  
<!ATTLIST score algorithm (default | count) "default">  
```

All tags and attributes values are case-sensitive.

**See Also:** Chapter 3, "Oracle Text CONTAINS Query Operators" for more information about the operators in query expressions.

**label**

Optionally, specify the label that identifies the score generated by the `CONTAINS` operator.

**Returns**

For each row selected, `CONTAINS` returns a number between 0 and 100 that indicates how relevant the document row is to the query. The number 0 means that Oracle Text found no matches in the row.

**Note:** You must use the `SCORE` operator with a label to obtain this number.
Example

The following example searches for all documents in the text column that contain the word oracle. The score for each row is selected with the SCORE operator using a label of 1:

```
SELECT SCORE(1), title from newsindex
  WHERE CONTAINS(text, 'oracle', 1) > 0;
```

The CONTAINS operator must be followed by an expression such as > 0, which specifies that the score value calculated must be greater than zero for the row to be selected.

When the SCORE operator is called (for example, in a SELECT clause), the CONTAINS clause must reference the score label value as in the following example:

```
SELECT SCORE(1), title from newsindex
  WHERE CONTAINS(text, 'oracle', 1) > 0 ORDER BY SCORE(1) DESC;
```

The following example specifies that the query be parsed using the CATSEARCH grammar:

```
SELECT id FROM test WHERE CONTAINS (text, '<query>
  <textquery lang="ENGLISH" grammar="CATSEARCH">cheap pokemon</textquery>
  <score datatype="INTEGER"/>
</query>' ) > 0;
```

Grammar Template Example

The following example shows how to use the CTXCAT grammar in a CONTAINS query. The example creates a CTXCAT and a CONTEXT index on the same table, and compares the query results:

```
PROMPT create context and ctxcat indexes both with theme indexing on
PROMPT
create index tdrbqcq101x on test(text) indextype is ctxsys.context
parameters ('lexer theme_lexer');
create index tdrbqcq101cx on test(text) indextype is ctxsys.ctxcat
parameters ('lexer theme_lexer');

PROMPT *****  San Diego             ***********
PROMPT *****  CONTEXT grammar       ***********
PROMPT ** should be interpreted as phrase query **
select pk||' ==> '||text from test
where contains(text,'San Diego')>0
order by pk;

PROMPT *****  San Diego             ***********
PROMPT *****  CTXCAT grammar        ***********
PROMPT ** should be interpreted as AND query ***
select pk||' ==> '||text from test
where contains(text, '<query>
  <textquerygrammar="CTXCAT">San Diego</textquery>
  <score datatype="integer"/>
</query>' )>0
order by pk;
```
PROMPT ***** Hitlist from CTXCAT index ************
select pk || ' ==> ' || text from test
where catsearch(text, 'San Diego', '') > 0
order by pk;

Alternate Scoring Query Template Example
The following query template adds price SDATA section (or SDATA filter-by-column) value into the score returned by the query and uses it as the final score.

<query>
  <textquery grammar="CONTEXT" lang="english">
    DEFINESCORE(dog, RELEVANCE) and cat
  </textquery>
  <score algorithm="COUNT" normalization_expr = "doc_score+SDATA(price)"/>
</query>

Query Relaxation Template Example
The following query template defines a query relaxation sequence. The query of black pen is entered in sequence as black pen then black NEAR pen then black AND pen then black ACCUM pen. Query hits are returned in this sequence with no duplication as long as the application needs results.

select id from docs where CONTAINS (text, 'black pen')>0;

Query relaxation is most effective when your application needs the top n hits to a query, which you can obtain with the DOMAIN_INDEX_SORT or FIRST_ROWS hint, which is being deprecated, in a PL/SQL cursor.

Query Rewrite Example
The following template defines a query rewrite sequence. The query of kukui nut is rewritten as follows:

{kukui} {nut}
{kukui} ; {nut}
{kukui} AND {nut}
{kukui} ACCUM {nut}

select id from docs where CONTAINS (text, 'kukui nut')>0;

<query>
  <textquery lang="ENGLISH" grammar="CONTEXT">
    <progression>
      <seq>black pen</seq>
      <seq>black NEAR pen</seq>
      <seq>black AND pen</seq>
      <seq>black ACCUM pen</seq>
    </progression>
  </textquery>
  <score datatype="INTEGER" algorithm="COUNT"/>
</query>
<syntax>
<score datatype='INTEGER' algorithm='COUNT'/>
</query> > 0;
</syntax>

Notes

**Querying Multi-Language Tables**
With the multi-lexer preference, you can create indexes from multi-language tables. At
query time, the multi-lexer examines the session's language setting and uses the
sub-lexer preference for that language to parse the query. If the language setting is not
mapped, then the default lexer is used.

When the language setting is mapped, the query is parsed and run as usual. The index
contains tokens from multiple languages, so such a query can return documents in
several languages.

To limit your query to returning document of a given language, use a structured
clause on the language column.

**Query Performance Limitation with a Partitioned Index**
Oracle Text supports the CONTEXT indexing and querying of a partitioned text table.

However, for optimal performance when querying a partitioned table with an ORDER
BY SCORE clause, query the partition. If you query the entire table and use an ORDER
BY SCORE clause, the query might not perform optimally unless you include a range
predicate that can limit the query to a single partition.

For example, the following statement queries the partition p_tab4 partition directly:

```sql
select * from part_tab partition (p_tab4) where contains(b,'oracle') > 0 ORDER BY
SCORE DESC;
```

Related Topics

- Syntax for CONTEXT Index Type in this chapter
- Chapter 3, "Oracle Text CONTAINS Query Operators"
- "The CONTEXT Grammar" topic in Oracle Text Application Developer’s Guide
- SCORE in this chapter
CREATE INDEX

This section describes the CREATE INDEX statement as it pertains to creating an Oracle Text domain index and composite domain index.

See Also: "Oracle Database SQL Language Reference" for a complete description of the CREATE INDEX statement

Purpose

Use CREATE INDEX to create an Oracle Text index. An Oracle Text index is an Oracle Database domain index or composite domain index of type CONTEXT, CTXCAT, CTXRULE, or CTXXPATH. A domain index is an application-specific index. A composite domain index (CDI) is an Oracle Text index that not only indexes and processes a specified text column, but also indexes and processes FILTER BY and ORDER BY structured columns, which are specified during index creation.

You must create an appropriate Oracle Text index to enter CONTAINS, CATSEARCH, or MATCHES queries.

You cannot create an Oracle Text index on an index-organized table.

You can create the following types of Oracle Text indexes:

CONTEXT

A CONTEXT index is the basic type of Oracle Text index. This is an index on a text column. A CONTEXT index is useful when your source text consists of many large, coherent documents. Query this index with the CONTAINS operator in the WHERE clause of a SELECT statement. This index requires manual synchronization after DML. See "Syntax for CONTEXT Index Type" on page 1-37.

CTXCAT

The CTXCAT type of index is a combined index on a text column and one or more other columns. CTXCAT is typically used to index small documents or text fragments, such as item names, prices and descriptions found in catalogs. Query this index with the CATSEARCH operator in the WHERE clause of a SELECT statement. This type of index is optimized for mixed queries. This index is transactional, automatically updating itself with DML to the base table. See "Syntax for CTXCAT Index Type" on page 1-52.

CTXRULE

A CTXRULE index is used to build a document classification application. The CTXRULE index is an index created on a table of queries or a column containing a set of queries, where the queries serve as rules to define the classification criteria. Query this index with the MATCHES operator in the WHERE clause of a SELECT statement. See "Syntax for CTXRULE Index Type" on page 1-55.

CTXXPATH

Create this index when you need to speed up existsNode() queries on an XMLType column. See "Syntax for CTXXPATH Index Type" on page 1-56.
**Required Privileges**
You do not need the `CTXAPP` role to create an Oracle Text index. If you have Oracle Database grants to create a b-tree index on the text column, you have sufficient permission to create a text index. The issuing owner, table owner, and index owner can all be different users, which is consistent with Oracle standards for creating regular B-tree indexes.

**Syntax for CONTEXT Index Type**
Use a `CONTEXT` index to create an index on a text column. Query this index with the `CONTAINS` operator in the `WHERE` clause of a `SELECT` statement. This index requires manual synchronization after DML.

```
CREATE INDEX [schema.]index ON [schema.]table(txt_column)
  INDEXTYPE IS ctxsys.context [ONLINE]
  [FILTER BY filter_column[, filter_column]...]
  [ORDER BY oby_column[desc|asc][, oby_column[desc|asc]]...]
  [LOCAL [(PARTITION [partition] [PARAMETERS('paramstring')])
            [, PARTITION [partition] [PARAMETERS('paramstring')]])]
        [PARAMETERS(paramstring)] [PARALLEL n] [UNUSABLE];
```

**[schema.]index**
Specify the name of the Text index to create.

**[schema.]table(txt_column)**
Specify the name of the table and column to index. `txt_column` is the name of the domain index column on which the `CONTAINS()` operator will be invoked.

Your table can optionally contain a primary key if you prefer to identify your rows as such when you use procedures in `CTX_DOC`. When your table has no primary key, document services identifies your documents by `ROWID`.

The column that you specify must be one of the following types: `CHAR`, `VARCHAR`, `VARCHAR2`, `BLOB`, `CLOB`, `BFILE`, `XMLType`, or `URIType`.

The table that you specify can be a partitioned table. If you do not specify the `LOCAL` clause, then a global, nonpartitioned index is created.

`DATE`, `NUMBER`, and nested table columns cannot be indexed. Object columns also cannot be indexed, but their attributes can be indexed, provided that they are atomic data types.

Attempting to create an index on a Virtual Private Database (VPD) protected table will fail unless one of the following criteria is true:

- The VPD policy is created such that it does not apply to `INDEX` statement type.
- The policy function returns a NULL predicate for the current user.
- The user (or index owner) is `SYS`.
- The user has the `EXEMPT ACCESS POLICY` privilege.

Indexes on multiple columns are not supported with the `CONTEXT` index type. You must specify only one column in the column list.

**Note:** With the `CTXCAT` index type, you can create indexes on text and structured columns. See "Syntax for CTXCAT Index Type" on page 1-52 in this chapter.
ONLINE
Creates the index while enabling DML inserts/uploads/deletes on the base table.

During indexing, Oracle Text enqueues DML requests in a pending queue. At the end of the index creation, Oracle Text locks the base table. During this time DML is blocked. You must synchronize the index in order for DML changes to be available.

Limitations
The following limitations apply to using ONLINES:

- At the very beginning or very end of the ONLINE process, DML might fail.
- ONLINE is supported for CONTEXT indexes only.

FILTER BY filter_column
This is the structured indexed column on which a range or equality predicate in the WHERE clause of a mixed query will operate. You can specify one or more structured columns for filter_column, on which the relational predicates are expected to be specified along with the CONTAINS() predicate in a query.

The cost-based optimizer (CBO) will consider pushing down the structured predicates on these FILTER BY columns with the following relational operators: <, <=, =, >, >=, between, and LIKE (for VARCHAR2).

These columns can only be of CHAR, NUMBER, DATE, VARCHAR2, or RAW type. Additionally, VARCHAR2 and RAW types are only supported if the maximum length is specified and is limited to no more than 249. ADT attributes of supported types (CHAR, NUMBER, DATE, VARCHAR2, or RAW) are also allowed. An error is raised for all other datatypes. Expressions, for example, func(cola), and virtual columns are not allowed.

txt_column is allowed in the FILTER BY column list.

DML operations on FILTER BY columns are always transactional.

ORDER BY oby_column
This is the structured indexed column on which a structured ORDER BY mixed query will be based. A list of structured oby_columns can be specified in the ORDER BY clause of a CONTAINS() query.

These columns can only be of CHAR, NUMBER, DATE, VARCHAR2, or RAW type. Additionally, VARCHAR2 and RAW types are only supported if the maximum length is specified and is limited to no more than 249. Expressions, for example, func(cola), and virtual columns are not allowed.

The order of the specified columns matters. The cost based optimizer will consider pushing the sort into the composite domain index only if the ORDER BY clause in the text query contains:

- entire ordered ORDER BY columns declared by the ORDER BY clause during CREATE INDEX,

CBO will consider pushing the sort into the CDI only if the ORDER BY clause in the text query contains:
CREATE INDEX

- entire ordered ORDER BY columns declared by the ORDER BY clause during CREATE INDEX,
- only the prefix of the ordered ORDER BY columns declared by the ORDER BY clause during CREATE INDEX,
- score followed by the prefix of the ordered ORDER BY columns declared by the ORDER BY clause during CREATE INDEX,
- or score following the prefix of the ordered ORDER BY columns declared by the ORDER BY clause during CREATE INDEX.

The following example illustrates CBO behavior with regard to ORDER BY columns:

CREATE INDEX foox ON foo(D) INDEXTYPE IS CTXSYS.CONTEXT
FILTER BY B, C
ORDER BY A, B desc;

Consider the following query:

SELECT A, SCORE(1) FROM foo WHERE CONTAINS(D, 'oracle',1)>0
AND C>100 ORDER BY col_list;

Note: If you set NLS_SORT or NLS_COMP parameters (that is, alter session set NLS_SORT = <some lang>), then CBO will not push the Sort or related structured predicate into the CDI. This behavior is consistent with regular btree indexes.

CBO will consider pushing the sort into CDI if col_list has the following values:

A
A, B desc
SCORE(1), A
SCORE(1), A, B desc
A, SCORE(1)
A, B desc, SCORE(1)

CBO will NOT consider to push the sort into CDI if col_list has the following values:

B
B, A
SCORE(1), B
B, SCORE(1)
A, B, C
A, B asc (or simply A, B)
(or simply A, B)
- score followed by the prefix of the ordered ORDER BY columns declared by the ORDER BY clause during CREATE INDEX,
- or score following the prefix of the ordered ORDER BY columns declared by the ORDER BY clause during CREATE INDEX.

Expressions, for example, func(cola), are not allowed.

txt_column appearing in the ORDER BY column list is allowed.

DML operations on ORDER BY columns are always transactional.
Limitations
The following limitations apply to FILTER BY and ORDER BY:

- A structured column is allowed in FILTER BY and ORDER BY clauses. However, a column that is mapped to MDATA in a FILTER BY clause cannot also appear in the ORDER BY clause. An error will be raised in this case.

- The maximum length for CHAR, VARCHAR2 and RAW columns cannot be greater than 249. Additionally, if the VARCHAR2 or RAW column is mapped to an MDATA column, then the specified maximum length cannot exceed 64 and 32 bytes, respectively. (Note that MDATA does not support CHAR datatypes. If a FILTER BY column of CHAR datatype is mapped to an MDATA section, then an error will be raised during CREATE INDEX.)

- The sum of number of FILTER BY and ORDER BY columns cannot be greater than 32.

Note:
- As with concatenated btree indexes or bitmap indexes, performance degradation may occur in DML as the number of FILTER BY and ORDER BY columns increases.

- Mapping a FILTER BY column to MDATA is not recommended if the FILTER BY column contains sequential values or has very high cardinality. Doing so can result in very long and narrow $I table and reduced $X performance. An example is a column of type DATE. For columns of this type, mapping to SDATA is recommended.

Note: An index table with the name DR$indextable$S is created to store FILTER BY and ORDER BY columns that are mapped to SDATA sections. If nothing is mapped to an SDATA section, then the $S table will not be created.

$S table contains the following columns:

- SDATA_ID number is the internal SDATA section id.
- SDATA_LAST number, the last docid, which is analogous to token_last.
- SDATA_DATA RAW(2000), the compressed SDATA values. Note that if $S is created on a tablespace with 4K db block size, then it will be defined as RAW(1500).

Restriction: For performance reasons, $S table must be created on a tablespace with db block size >= 4K without overflow segment and without PCTTHRESHOLD clause. If $S is created on a tablespace with db block size < 4K, or is created with an overflow segment or with PCTTHRESHOLD clause, then appropriate errors will be raised during CREATE INDEX.

Restriction on Exporting and Importing text tables with Composite Domain Index created with FILTER BY and/or ORDER BY clauses:
Regular `exp` and `imp` will not support exporting and importing of Composite Domain Index. Doing so will lead to the following error: EXP-00113: Feature Composite Domain Index is unsupported.

To export a text table with Composite Domain Index, you must use Data Pump Export and Import utilities (invoked with the `expdp` and `impdp` commands, respectively), or `DBMS_DATAPUMP` PL/SQL package.

**See Also:** `ADD_SDATA_COLUMN` in Chapter 7, "CTX_DDL Package"

Limitations of using `ALTER_INDEX` and `ALTER_TABLE` with `FILTER BY` and `ORDER BY` columns of the Composite Domain Index, which are imposed by Extensible Indexing Framework in Oracle Database:

(These limitations are imposed by Extensible Indexing Framework in Oracle Database.)

- Using `ALTER_INDEX` to add or drop `FILTER BY` and `ORDER BY` columns is currently not supported. You must re-create the index to add or drop `FILTER BY` or `ORDER BY` columns.
- To use `ALTER_TABLE MODIFY COLUMN` to modify the datatype of a column that has the composite domain index built on it, you must first drop the composite domain index before modifying the column.
- To use `ALTER_TABLE DROP COLUMN` to drop a column that is part of the composite domain index, you must first drop the composite domain index before dropping the index column.

The following limitations apply to `FILTER BY` and `ORDER BY` when used with PL/SQL packages:

- Mapping `FILTER BY` columns to sections is optional. If section mapping does not exist for a `FILTER BY` column, then it is mapped to an `SDATA` section by default. The section name assumes the name of the `FILTER BY` column.
- If a section group is not specified during `CREATE_INDEX` of a composite domain index, then system default section group settings are used. An `SDATA` section is created for each of the `FILTER BY` and `ORDER BY` columns.

---

**Note:** Since section name does not allow certain special characters and is case insensitive, if the column name is case sensitive or contains special characters, then an error will be raised. To work around this problem, you must map the column to an `MDATA` or `SDATA` section before creating the index. See `CTX_DDL.ADD_MDATA_COLUMN` or `CTX_DDL.ADD_SDATA_COLUMN`.

---

- An error is raised if a column that is mapped to `MDATA` also appears in the `ORDER BY` column clause.
- Column section names are unique to their section group. That is, you cannot have an `MDATA` column section named `FOO` if you already have an `MDATA` column section named `FOO`. Nor can you have a field section named `FOO` if you already have an `SDATA` column section named `FOO`. This is true whether it is implicitly created (by `CREATE_INDEX` for `FILTER BY` or `ORDER BY` clauses) or explicitly created (by `CTX_DDL.ADD_SDATA_COLUMN`).

---
- One section name can only be mapped to one `FILTER BY` column, and vice versa. Mapping a section to more than one column, or mapping a column to more than one section is not allowed.

- Column sections can be added to any type of section group, including the NULL section group.

- If a section group with sections added by `CTX_DDL.ADD_MDATA_COLUMN` or `CTX_DDL.ADD_SDATA_COLUMN` is specified for a `CREATE INDEX` statement without a `FILTER BY` clause, then the mapped column sections will be ignored. However, the index will still get created without those column sections. The same is true for a `FILTER BY` clause that does not contain mapped column(s) in the specified section group.

**See Also:** `CTX_DDL.ADD_SDATA_COLUMN`

**LOCAL** `{PARTITION [partition] [PARAMETERS('paramstring')]}`  
Specify `LOCAL` to create a local partitioned context index on a partitioned table. The partitioned table must be partitioned by range. Hash, composite and list partitions are not supported.

You can specify the list of index partition names with `partition_name`. If you do not specify a partition name, then the system assigns one. The order of the index partition list must correspond to the table partition by order.

The `PARAMETERS` clause associated with each partition specifies the parameters string specific to that partition. You can only specify `sync (manual | every | on commit)`, `memory` and `storage` for each index partition.

The `PARAMETERS` clause also supports `POPULATE` and `NOPOPULATE`. See `POPULATE` | `NOPOPULATE` on page 1-46.

Query the views `CTX_INDEX_PARTITIONS` or `CTX_USER_INDEX_PARTITIONS` to find out index partition information, such as index partition name, and index partition status.

**See Also:** "Creating a Local Partitioned Index” on page 1-49

**Query Performance Limitation with Partitioned Index**  
For optimal performance when querying a partitioned index with an `ORDER BY SCORE` clause, query the partition. If you query the entire table and use an `ORDER BY SCORE` clause, the query might not perform optimally unless you include a range predicate that can limit the query to the fewest number of partitions, which is optimally a single partition.

**See Also:** "Query Performance Limitation with a Partitioned Index” on page 1-35 in this chapter under CONTAINS

**PARALLEL** `{n}`  
Optionally specify with `n` the parallel degree for parallel indexing. The actual degree of parallelism might be smaller depending on your resources.

You can use this parameter on nonpartitioned tables. However, creating a nonpartitioned index in parallel does not turn on parallel query processing.

Parallel indexing is supported for creating a local partitioned index.
**Performance**
Parallel indexing can speed up indexing when you have large amounts of data to index and when your operating system supports multiple CPUs.

**Note:** Using PARALLEL to create a local partitioned index enables parallel queries. (Creating a nonpartitioned index in parallel does not turn on parallel query processing.)

Parallel querying degrades query throughput especially on heavily loaded systems. Because of this, Oracle recommends that you disable parallel querying after creating a local index. To do so, use ALTER INDEX NOPARALLEL.

For more information on parallel querying, see the "Performance Tuning" chapter in *Oracle Text Application Developer’s Guide*.

**Limitations**
The following limitations apply to using PARALLEL:

- Parallel indexing is supported only for CONTEXT index type.

**UNUSABLE**
Create an unusable index. This creates index metadata only and exits immediately.

You might create an unusable index when you need to create a local partitioned index in parallel.

**See Also:** "Creating a Local Partitioned Index in Parallel" on page 1-50

**PARAMETERS**(*paramstring*)
Optionally specify indexing parameters in *paramstring*. You can specify preferences owned by another user using the user.preference notation.

The syntax for *paramstring* is as follows:

```
paramstring =
  '[DATASTORE datastore_pref]
  [FILTER filter_pref]
  [CHARSET COLUMN charset_column_name]
  [FORMAT COLUMN format_column_name]
  [LEXER lexer_pref]
  [LANGUAGE COLUMN language_column_name]
  [WORDLIST wordlist_pref]
  [STORAGE storage_pref]
  [STOPLIST stoplist]
  [SECTION GROUP section_group]
  [MEMORY memsize]
```

**See Also:** "Creating a Local Partitioned Index in Parallel" on page 1-50
CREATE INDEX

[POPULATE | NOPOPULATE]
[SYNC (MANUAL | EVERY "interval-string" | ON COMMIT) | TRANSACTIONAL]

Create datastore, filter, lexer, wordlist, and storage preferences with CTX_DDL.CREATE_PREFERENCE and then specify them in the paramstring.

Note: When you specify no paramstring, Oracle Text uses the system defaults.
For more information about these defaults, see "Default Index Parameters" on page 2-88 in Chapter 2.

DATASTORE datastore_pref
Specify the name of your datastore preference. Use the datastore preference to specify where your text is stored. See Datastore Types on page 2-2 in Chapter 2, "Oracle Text Indexing Elements".

FILTER filter_pref
Specify the name of your filter preference. Use the filter preference to specify how to filter formatted documents to plain text or HTML. See Filter Types on page 2-16 in Chapter 2, "Oracle Text Indexing Elements".

CHARSET COLUMN charset_column_name
Specify the name of the character set column. This column must be in the same table as the text column, and it must be of type CHAR, VARCHAR, or VARCHAR2. Use this column to specify the document character set for conversion to the database character set. The value is case insensitive. You must specify a globalization support character set string, such as JA16EUC.

When the document is plain text or HTML, the AUTO_FILTER and CHARSET filter use this column to convert the document character set to the database character set for indexing.

Use this column when you have plain text or HTML documents with different character sets or in a character set different from the database character set.

NLS_LENGTH_SEMANTICS = CHAR is not supported at the database level. This parameter is supported for the following columns:

- The CHARSET COLUMN, for example:
  
  VARCHAR2 <size> CHAR
  CHAR <size> CHAR

- Creating an index on a VARCHAR2 and CHAR column

- Using a VARCHAR2 and CHAR column for FILTER BY and ORDER BY clause of CREATE INDEX

- FORMAT COLUMN
FORMAT COLUMN `format_column_name`

Specify the name of the format column. The format column must be in the same table as the text column and it must be `CHAR`, `VARCHAR`, or `VARCHAR2` type.

FORMAT COLUMN determines how a document is filtered, or, in the case of the IGNORE value, if it is to be indexed.

The AUTO_FILTER uses the format column when filtering documents. Use this column with heterogeneous document sets to optionally bypass filtering for plain text or HTML documents.

In the format column, you can specify one of the following options:

- **TEXT**
- **BINARY**
- **IGNORE**

TEXT indicates that the document is either plain text or HTML. When TEXT is specified, the document is not filtered, but may have the character set converted.

BINARY indicates that the document is a format supported by the AUTO_FILTER object other than plain text or HTML, for example PDF. BINARY is the default, if the format column entry cannot be mapped.

IGNORE indicates that the row is to be ignored during indexing. Use this value when you need to bypass rows that contain data incompatible with text indexing such as image data, or rows in languages that you do not want to process. The difference between documents with TEXT and IGNORE format column types is that the former are indexed but ignored by the filter, while the latter are not indexed at all. Thus, IGNORE can be used with any filter type.

**Note:** Documents are not marked for re-indexing when only the format column changes. The indexed column must be updated to flag the re-index.

LEXER `lexer_pref`

Specify the name of your lexer or multi-lexer preference. Use the lexer preference to identify the language of your text and how text is tokenized for indexing. See Lexer Types on page 2-30 in Chapter 2, "Oracle Text Indexing Elements".

LANGUAGE COLUMN `language_column_name`

Specify the name of the language column when using a multi-lexer preference. See MULTI_LEXER on page 2-48 in Chapter 2, "Oracle Text Indexing Elements".

This column must exist in the base table. It cannot be the same column as the indexed column. Only the first 30 bytes of the language column is examined for language identification.
WORDLIST wordlist_pref
Specify the name of your wordlist preference. Use the wordlist preference to enable features such as fuzzy, stemming, and prefix indexing for better wildcard searching. See Wordlist Type on page 2-68 in Chapter 2, "Oracle Text Indexing Elements".

STORAGE storage_pref
Specify the name of your storage preference for the Text index. Use the storage preference to specify how the index tables are stored. See Storage Types on page 2-75 in Chapter 2, "Oracle Text Indexing Elements".

STOPLIST stoplist
Specify the name of your stoplist. Use stoplist to identify words that are not to be indexed. See CTX_DDL.CREATE_STOPLIST in Chapter 7, "CTX_DDL Package".

SECTION GROUP section_group
Specify the name of your section group. Use section groups to create searchable sections in structured documents. See CTX_DDL.CREATE_SECTION_GROUP in Chapter 7, "CTX_DDL Package".

MEMORY memsize
Specify the amount of run-time memory to use for indexing. The syntax for memsize is as follows:

\[
\text{memsize} = \text{number}[K|M|G]
\]

where K stands for kilobytes, M stands for megabytes, and G stands for gigabytes.

The value you specify for memsize must be between 1M and the value of MAX_INDEX_MEMORY in the CTX_PARAMETERS view. To specify a memory size larger than the MAX_INDEX_MEMORY, you must reset this parameter with CTX_ADM.SET_PARAMETER to be larger than or equal to memsize.

The default is the value specified for DEFAULT_INDEX_MEMORY in CTX_PARAMETERS.

The memsize parameter specifies the amount of memory Oracle Text uses for indexing before flushing the index to disk. Specifying a large amount memory improves indexing performance because there are fewer I/O operations and improves query performance and maintenance, because there is less fragmentation.

Specifying smaller amounts of memory increases disk I/O and index fragmentation, but might be useful when run-time memory is scarce.

POPULATE | NOPOPULATE
Specify NOPOPULATE to create an empty index. The default is POPULATE.

Note: POPULATE | NOPOPULATE is the only option whose default value cannot be set with CTX_ADM.SET_PARAMETER.

This option is not valid with CTXXPATH indexes.

Empty indexes are populated by updates or inserts to the base table. You might create an empty index when you need to create your index incrementally or to selectively
index documents in the base table. You might also create an empty index when you require only theme and Gist output from a document set.

SYNC (MANUAL | EVERY "interval-string" | ON COMMIT)
Specify SYNC for automatic synchronization of the CONTEXT index when there are inserts, updates or deletes to the base table. You can specify one of the following SYNC methods:

<table>
<thead>
<tr>
<th>SYNC type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MANUAL</td>
<td>No automatic synchronization. This is the default. You must manually synchronize the index with CTX_DDL.SYNC_INDEX.</td>
</tr>
<tr>
<td>EVERY &quot;interval-string&quot;</td>
<td>Automatically synchronize the index at a regular interval specified by the value of interval-string. interval-string takes the same syntax as that for scheduler jobs. Automatic synchronization using EVERY requires that the index creator have CREATE JOB privileges. Make sure that interval-string is set to a long enough period that any previous sync jobs will have completed; otherwise, the sync job may hang. interval-string must be enclosed in double quotes, and any single quote within interval-string must be escaped with another single quote. See “Enabling Automatic Index Synchronization” on page 1-49 for an example of automatic sync syntax.</td>
</tr>
<tr>
<td>ON COMMIT</td>
<td>Synchronize the index immediately after a commit. The commit does not return until the sync is complete. (Because the synchronization is performed as a separate transaction, there may be a period, usually small, when the data is committed but index changes are not.) The operation uses the memory specified with the memory parameter. Note that the sync operation has its own transaction context. If this operation fails, the data transaction still commits. Index synchronization errors are logged in the CTX_USER_INDEX_ERRORS view. See &quot;Viewing Index Errors&quot; under CREATE INDEX. See “Enabling Automatic Index Synchronization” on page 1-49 for an example of ON COMMIT syntax.</td>
</tr>
</tbody>
</table>

Each partition of a locally partitioned index can have its own type of sync (ON COMMIT, EVERY, or MANUAL). The type of sync specified in master parameter strings applies to all index partitions unless a partition specifies its own type.

With automatic (EVERY) synchronization, users can specify memory size and parallel synchronization. That syntax is:

... EVERY interval_string MEMORY mem_size PARALLEL paradegree ...

ON COMMIT synchronizations can only be run serially and must use the same memory size that was specified at index creation.

*See Also:* Oracle Database Administrator’s Guide for information on job scheduling

**TRANSACTIONAL**
Specify that documents can be searched immediately after they are inserted or updated. If a text index is created with TRANSACTIONAL enabled, then, in addition to
processing the synchronized rowids already in the index, the CONTAINS operator will process unsynchronized rowids as well. Oracle Text does in-memory indexing of unsynchronized rowids and processes the query against the in-memory index.

**TRANSACTIONAL** is an index-level parameter and does not apply at the partition level.

You must still synchronize your text indexes from time to time (with CTX_DDL.SYNC_INDEX) to bring pending rowids into the index. Query performance degrades as the number of unsynchronized rowids increases. For that reason, Oracle recommends setting up your index to use automatic synchronization with the EVERY parameter. (See **SYNC (MANUAL | EVERY "interval-string" | ON COMMIT)** on page 1-47.)

Transactional querying for indexes that have been created with the **TRANSACTIONAL** parameter can be turned on and off (for the duration of a user session) with the PL/SQL variable CTX_QUERY.disable_transactional_query. This is useful, for example, if you find that querying is slow due to the presence of too many pending rowids. Here is an example of setting this session variable:

```sql
exec ctx_query.disable_transactional_query := TRUE;
```

If the index uses **AUTO_FILTER**, queries involving unsynchronized rowids will require filtering of unsynchronized documents.

**CREATE INDEX: CONTEXT Index Examples**

The following sections give examples of creating a CONTEXT index.

**Creating CONTEXT Index Using Default Preferences**

The following example creates a CONTEXT index called myindex on the docs column in mytable. Default preferences are used.

```sql
CREATE INDEX myindex ON mytable(docs) INDEXTYPE IS ctxsys.context;
```

See Also:

- Oracle Text Application Developer’s Guide
- For more information about default settings, see "Default Index Parameters" on page 2-88 in Chapter 2

**Creating CONTEXT Index with Custom Preferences**

The following example creates a CONTEXT index called myindex on the docs column in mytable. The index is created with a custom lexer preference called my_lexer and a custom stoplist called my_stop.

This example also assumes that the preference and stoplist were previously created with CTX_DDL.CREATE_PREFERENCE for my_lexer, and CTX_DDL.CREATE_STOPLIST for my_stop. Default preferences are used for the unspecified preferences.

```sql
CREATE INDEX myindex ON mytable(docs) INDEXTYPE IS ctxsys.context
    PARAMETERS('LEXER my_lexer STOPLIST my_stop');
```

Any user can use any preference. To specify preferences that exist in another user's schema, add the user name to the preference name. The following example assumes that the preferences my_lexer and my_stop exist in the schema that belongs to user kenny:

```sql
CREATE INDEX myindex ON mytable(docs) INDEXTYPE IS ctxsys.context
    PARAMETERS('LEXER kenny.my_lexer STOPLIST kenny.my_stop');
```
Enabling Automatic Index Synchronization

You can create your index and specify that the index be synchronized at regular intervals for inserts, updates and deletes to the base table. To do so, create the index with the `SYNC (EVERY "interval-string")` parameter.

To use job scheduling, you must log in as a user who has DBA privileges and then grant CREATE JOB privileges.

The following example creates an index and schedules three synchronization jobs for three index partitions. The first partition uses ON COMMIT synchronization. The other two partitions are synchronized by jobs that are scheduled to be executed every Monday at 3 P.M.

```sql
CONNECT system/manager
GRANT CREATE JOB TO dr_test
CREATE INDEX tdrmauto02x ON tdrmauto02(text)
    INDEXTYPE IS CTXSYSCONTEXT local
    (PARTITION tdrm02x_i1 PARAMETERS('MEMORY 20m SYNC (ON COMMIT)'),
        PARTITION tdrm02x_i2,
        PARTITION tdrm02x_i3) PARAMETERS('SYNC (EVERY 'NEXT_DAY(TRUNC(SYSDATE), 'MONDAY') + 15/24')');
```

See the Oracle Database Administrator’s Guide for information on job scheduling syntax.

Creating CONTEXT Index with Multi-Lexer Preference

The multi-lexer decides which lexer to use for each row based on a language column. This is a character column in the table which stores the language of the document in the text column. For example, create the table `globaldoc` to hold documents of different languages:

```sql
CREATE TABLE globaldoc {
    doc_id NUMBER PRIMARY KEY,
    lang VARCHAR2(10),
    text CLOB
};
```

Assume that `global_lexer` is a multi-lexer preference you created. To index the `global_doc` table, specify the multi-lexer preference and the name of the language column as follows:

```sql
CREATE INDEX globalx ON globaldoc(text) INDEXTYPE IS ctxsys.context PARAMETERS ('LEXER global_lexer LANGUAGE COLUMN lang');
```

See Also: For more information about creating multi-lexer preferences, see "MULTI_LEXER" on page 2-48 in Chapter 2.

Creating a Local Partitioned Index

The following example creates a text table that is partitioned into three, populates it, and then creates a partitioned index.

```sql
PROMPT create partitioned table and populate it
CREATE TABLE part_tab (a int, b varchar2(40)) PARTITION BY RANGE(a)
(partition p_tab1 values less than (10),
    partition p_tab2 values less than (20),
    partition p_tab3 values less than (30));
```
CREATE INDEX part_idx on part_tab(b) INDEXTYPE IS CTXSYS.CONTEXT
LOCAL (partition p_idx1, partition p_idx2, partition p_idx3);

**Note:** The limit for the number of partitions in Oracle Text is the same as the maximum number of partitions per table in Oracle Database.

### Using FILTER BY and ORDER BY Clauses
The following example creates an index on table `docs` and orders the documents by author's publishing date.

First, create the table:

```sql
CREATE TABLE docs (
    docid NUMBER,
    pub_date DATE,
    author VARCHAR2(30),
    category VARCHAR2(30),
    document CLOB
);
```

Create the index with FILTER BY and ORDER BY clauses:

```sql
CREATE INDEX doc_idx on docs(document) indextype is ctxsys.context
FILTER BY category, author
ORDER BY pub_date desc, docid
PARAMETERS ('memory 500M');
```

### Parallel Indexing
Parallel indexing can improve index performance when you have multiple CPUs.

To create an index in parallel, use the `PARALLEL` clause with a parallel degree. This example uses a parallel degree of 3:

```sql
CREATE INDEX myindex ON mytab(pk) INDEXTYPE IS ctxsys.context PARALLEL 3;
```

### Creating a Local Partitioned Index in Parallel
Creating a local partitioned index in parallel can improve performance when you have multiple CPUs. With partitioned tables, you can divide the work. You can create a local partitioned index in parallel in two ways:

- Use the `PARALLEL` clause with the `LOCAL` clause in `CREATE INDEX`. In this case, the maximum parallel degree is limited to the number of partitions you have. See "Parallelism with CREATE INDEX" on page 1-51.

- Create an unusable index first, then run the `DBMS_PCLXUTIL.BUILD_PART_INDEX` utility. This method can result in a higher degree of parallelism, especially if you have more CPUs than partitions. See "Parallelism with DBMS_PCLXUTIL.BUILD_PART_INDEX" on page 1-51.

If you attempt to create a local partitioned index in parallel, and the attempt fails, you may see the following error message:

`ORA-29953: error in the execution of the ODCIIndexCreate routine for one or more of the index partitions`

To determine the specific reason why the index creation failed, query the `CTX_USER_INDEX_ERRORS` view.
Parallelism with CREATE INDEX

You can achieve local index parallelism by using the PARALLEL and LOCAL clauses in CREATE INDEX. In this case, the maximum parallel degree is limited to the number of partitions that you have.

The following example creates a table with three partitions, populates them, and then creates the local indexes in parallel with a degree of 2:

```sql
create table part_tab3(id number primary key, text varchar2(100))
partition by range(id)
(partition p1 values less than (1000),
 partition p2 values less than (2000),
 partition p3 values less than (3000));

begin
  for i in 0..2999
  loop
    insert into part_tab3 values (i,'oracle');
  end loop;
end;
/

create index part_tab3x on part_tab3(text)
indextype is ctxsys.context local (partition part_tabx1,
 partition part_tabx2,
 partition part_tabx3)
parallel 2;
```

Parallelism with DBMS_PCLUTIL.BUILD_PART_INDEX

You can achieve local index parallelism by first creating an unusable CONTEXT index, and then running the DBMS_PCLUTIL.BUILD_PART_INDEX utility. This method can result in a higher degree of parallelism, especially when you have more CPUs than partitions.

In this example, the base table has three partitions. We create a local partitioned unusable index first, then run DBMS_PCLUTIL.BUILD_PART_INDEX, which builds the 3 partitions in parallel (referred to as inter-partition parallelism). Also, inside each partition, index creation proceeds in parallel (called intra-partition parallelism) with a parallel degree of 2. Therefore, the total parallel degree is 6 (3 times 2).

```sql
create table part_tab3(id number primary key, text varchar2(100))
partition by range(id)
(partition p1 values less than (1000),
 partition p2 values less than (2000),
 partition p3 values less than (3000));

begin
  for i in 0..2999
  loop
    insert into part_tab3 values (i,'oracle');
  end loop;
end;
/

create index part_tab3x on part_tab3(text)
indextype is ctxsys.context local (partition part_tabx1,
 partition part_tabx2,
 partition part_tabx3)
unusable;
```
exec dbms_pclxutil.build_part_index(jobs_per_batch=>3,
    procs_per_job=>2,
    tab_name=>'PART_TAB3',
    idx_name=>'PART_TAB3X',
    force_opt=>TRUE);

Viewing Index Errors

After a CREATE INDEX or ALTER INDEX operation, you can view index errors with Oracle Text views. To view errors on your indexes, query the CTX_USER_INDEX_ERRORS view. To view errors on all indexes as CTXSYS, query the CTX_INDEX_ERRORS view.

For example, to view the most recent errors on your indexes, enter the following statements:

```
SELECT err_timestamp, err_text FROM ctx_user_index_errors ORDER BY err_timestamp DESC;
```

Deleting Index Errors

To clear the index error view, enter the following statement:

```
DELETE FROM ctx_user_index_errors;
```

Syntax for CTXCAT Index Type

The CTXCAT index is a combined index on a text column and one or more other columns. Query this index with the CATSEARCH operator in the WHERE clause of a SELECT statement. This type of index is optimized for mixed queries. This index is transactional, automatically updating itself with DML to the base table.

```
CREATE INDEX [schema.]index on [schema.]table(column) INDEXTYPE IS ctxsys.ctxcat
    [PARAMETERS
    ('[index set index_set]
    [lexer lexer_pref]
    [storage storage_pref]
    [stoplist stoplist]
    [section group sectiongroup_pref]
    [wordlist wordlist_pref]
    [memory memsize']);

[schema.]table(column)
```

Specify the name of the table and column to index.

The column that you specify when you create a CTXCAT index must be of type CHAR or VARCHAR2. No other types are supported for CTXCAT.

Attempting to create an index on a Virtual Private Database (VPD) protected table will fail unless one of the following options is true:

- The VPD policy is created such that it does not apply to INDEX statement type, which is the default
- The policy function returns a null predicate for the current user.
- The user (index owner) is SYS.
- The user has the EXEMPT ACCESS POLICY privilege.
Supported CTXCAT Preferences

index set `index_set`
Specify the index set preference to create the CTXCAT index. Index set preferences name the columns that make up your sub-indexes. Any column that is named in an index set column list cannot have a NULL value in any row of the base table, or else you get an error.

Always ensure that your columns have non-NULL values before and after indexing.

See “Creating a CTXCAT Index” on page 1-54.

Index Performance and Size Considerations
Although a CTXCAT index offers query performance benefits, creating this type of index has its costs. The time that it takes Oracle Text to create a CTXCAT index depends on the total size of the index.

The total size of a CTXCAT index is directly related to:
- Total text to be indexed
- Number of component indexes in the index set
- Number of columns in the base table that make up the component indexes

Having many component indexes in your index set also degrades DML performance because more indexes must be updated.

Because of these added costs in creating a CTXCAT index, you should carefully consider the query performance benefit that each component index gives your application before adding it to your index set.

See Also: Oracle Text Application Developer’s Guide for more information about creating CTXCAT indexes and the benefits

Other CTXCAT Preferences
When you create an index of type CTXCAT, you can use the supported index preferences listed in Table 1–6 in the parameters string.

<table>
<thead>
<tr>
<th>Preference Class</th>
<th>Supported Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datastore</td>
<td>This preference class is not supported for CTXCAT.</td>
</tr>
<tr>
<td>Filter</td>
<td>This preference class is not supported for CTXCAT.</td>
</tr>
<tr>
<td>Lexer</td>
<td>AUTO_LEXER</td>
</tr>
<tr>
<td></td>
<td>BASIC_LEXER (index_themes attribute not supported)</td>
</tr>
<tr>
<td></td>
<td>CHINESE_LEXER</td>
</tr>
<tr>
<td></td>
<td>CHINESE_VGRAM_LEXER</td>
</tr>
<tr>
<td></td>
<td>JAPANESE_LEXER</td>
</tr>
<tr>
<td></td>
<td>JAPANESE_VGRAM_LEXER</td>
</tr>
<tr>
<td></td>
<td>KOREAN_MORPH_LEXER</td>
</tr>
<tr>
<td>Wordlist</td>
<td>BASIC_WORDLIST</td>
</tr>
<tr>
<td>Storage</td>
<td>BASIC_STORAGE</td>
</tr>
<tr>
<td>Stoplist</td>
<td>Supports single language stoplists only (BASIC_STOPLIST type).</td>
</tr>
<tr>
<td>Section Group</td>
<td>Only Field Section is supported for CTXCAT.</td>
</tr>
</tbody>
</table>
Unsupported Preferences and Parameters

When you create a CTXCAT index, you cannot specify datastore and filter preferences. For section group preferences, only the field section preference is supported. You also cannot specify language, format, or charset columns as with a CONTEXT index.

Creating a CTXCAT Index

This section gives a brief example for creating a CTXCAT index. For a more complete example, see Oracle Text Application Developer's Guide.

Consider a table called AUCTION with the following schema:

```
create table auction(
  item_id number,
  title varchar2(100),
  category_id number,
  price number,
  bid_close date);
```

Assume that queries on the table involve a mandatory text query clause and optional structured conditions on price. Results must be sorted based on bid_close. This means that we need an index to support good response time for the structured and sorting criteria.

You can create a catalog index to support the different types of structured queries a user might enter. For structured queries, a CTXCAT index improves query performance over a context index.

To create the indexes, first, create the index set preference, next, optionally, add the storage preference, and, finally, add the required indexes to it:

```
begin
  ctx_ddl.create_index_set('auction_iset');
  ctx_ddl.add_index('auction_iset','bid_close');
  ctx_ddl.add_index('auction_iset','price, bid_close');
end;
```

Optionally, create the storage preference:

```
begin
  ctx_ddl.create_preference('auction_st_pref', 'BASIC_STORAGE');
  ctx_ddl.set_attribute('auction_st_pref', 'I_TABLE_CLAUSE',
    'tablespace TEXT storage (initial 5M)');
  ctx_ddl.set_attribute('auction_st_pref', 'I_ROWID_INDEX_CLAUSE',
    'tablespace TEXT storage (initial 5M)');
  ctx_ddl.set_attribute('auction_st_pref', 'I_INDEX_CLAUSE',
    'tablespace TEXT storage (initial 5M) compress 2');
end;
/
```

Then, create the CTXCAT index with CREATE INDEX as follows:

```
cREATE INDEX auction_titlex ON AUCTION(title) indextype is CTXSYS.CTXCAT
parameters ('index set auction_iset storage auction_st_pref');
```

Querying a CTXCAT Index

To query the title column for the word *pokemon*, enter regular and mixed queries as follows:

```
select * from AUCTION where CATSEARCH(title, 'pokemon',NULL)> 0;
select * from AUCTION where CATSEARCH(title, 'pokemon', 'price < 50 order by
```
Syntax for CTXRULE Index Type

CTXRULE is an index on a column containing a set of queries. Query this index with the MATCHES operator in the WHERE clause of a SELECT statement.

```
CREATE INDEX [schema.]index on [schema.]table(rule_col) INDEXTYPE IS ctxsys.ctxrule
[PARAMETERS ('[lexer lexer_pref] [storage storage_pref]
[section group section_pref] [wordlist wordlist_pref]
[classifier classifier_pref]');
[PARALLEL n]);
```

**[schema.]table(column)**
Specify the name of the table and rule column to index. The rules can be query compatible strings, query template strings, or binary support vector machine rules.

The column you specify when you create a CTXRULE index must be VARCHAR2, CLOB or BLOB. No other types are supported for CTXRULE.

Attempting to create an index on a Virtual Private Database (VPD) protected table will fail unless one of the following is true:

- The VPD policy does not have the INDEX statement type turned on (which is the default)
- The policy function returns a null predicate for the current user.
- The user (index owner) is SYS.
- The user has the EXEMPT ACCESS POLICY privilege.

**lexer_pref**
Specify the lexer preference to be used for processing queries and later for the documents to be classified with the MATCHES function.

With both classifiers SVN_CLASSIFIER and RULE_CLASSIFIER, you can use the BASIC_LEXER, CHINESE_LEXER, JAPANESE_LEXER, or KOREAN_MORPH_LEXER lexer. (See "Classifier Types" on page 2-79 and "Lexer Types" on page 2-30.)

For processing queries, these lexers support the following operators: ABOUT, STEM, AND, NEAR, NOT, OR, and WITHIN.

The thesaural operators (BT*, NT*, PT, RT, SYN, TR, TRSYS, TT, and so on) are supported. However, these operators are expanded using a snapshot of the thesaurus at index time, not when the MATCHES function is entered. This means that if you change your thesaurus after you index, you must re-index your query set.

**storage_pref**
Specify the storage preference for the index on the queries. Use the storage preference to specify how the index tables are stored. See "Storage Types" on page 2-75 in Chapter 2, "Oracle Text Indexing Elements".

**section group**
Specify the section group. This parameter does not affect the queries. It applies to sections in the documents to be classified. The following section groups are supported for the CTXRULE index type:

- BASIC_SECTION_GROUP

See Also: Oracle Text Application Developer’s Guide for a complete CTXCAT example.
CREATE INDEX

- HTML_SECTION_GROUP
- XML_SECTION_GROUP
- AUTO_SECTION_GROUP

See "Section Group Types" on page 2-77 in Chapter 2, “Oracle Text Indexing Elements”. CTXRULE does not support special sections. It also does not support NDATA sections.

**wordlist_pref**
Specify the wordlist preferences. This is used to enable stemming operations on query terms. See Wordlist Type on page 2-68 in Chapter 2, "Oracle Text Indexing Elements".

**classifier_pref**
Specify the classifier preference. See "Classifier Types" on page 2-79 in Chapter 2, "Oracle Text Indexing Elements". You must use the same preference name you specify with CTX_CLS.TRAIN.

**Example for Creating a CTXRULE Index**

See the Oracle Text Application Developer’s Guide for a complete example of using the CTXRULE index type in a document routing application.

**Syntax for CTXXPATH Index Type**

This indextype if provided only for backward compatibility. Create a CTXXPATH index when you need to speed up existsNode() queries on an XMLType column.

```sql
CREATE INDEX [schema.]index on [schema.]table(XMLType column) INDEXTYPE IS ctxsys.CTXXPATH
[PARAMETERS ('[storage storage_pref]
 [memory memsize]')];
```

**[schema.]table(column)**
Specify the name of the table and column to index.

The column you specify when you create a CTXXPATH index must be XMLType. No other types are supported for CTXXPATH.

**storage_pref**
Specify the storage preference for the index on the queries. Use the storage preference to specify how the index tables are stored. See "Storage Types" on page 2-75 in Chapter 2, "Oracle Text Indexing Elements".

**memory memsize**
Specify the amount of run-time memory to use for indexing. The syntax for memsize is as follows:

```
memsize = number[M|G|K]
```

where M stands for megabytes, G stands for gigabytes, and K stands for kilobytes.

The value you specify for memsize must be between 1M and the value of MAX_INDEX_MEMORY in the CTX_PARAMETERS view. To specify a memory size larger than the MAX_INDEX_MEMORY, you must reset this parameter with CTX_ADM.SET_PARAMETER to be larger than or equal to memsize.

The default is the value specified for DEFAULT_INDEX_MEMORY in CTX_PARAMETERS.
**CTXXPath Examples**

Index creation on an XMLType column:

```
CREATE INDEX xml_index ON xml_tab(col_xml) indextype is ctxsys.CTXXPath;
```

or

```
CREATE INDEX xml_index ON xml_tab(col_xml) indextype is ctxsys.CTXXPath
PARAMETERS('storage my_storage memory 40M');
```

Querying the table with `existsNode`:

```
select xml_id from xml_tab x where 
x.col_xml.existsnode('/book/chapter[@title="XML"]') > 0;
```

**See Also:** Oracle XML DB Developer’s Guide for information on using the CTXXPath index type.

**Related Topics**

CTX_DDL.CREATE_PREFERENCE in Chapter 7, "CTX_DDL Package".

CTX_DDL.CREATE_STOPLIST in Chapter 7, "CTX_DDL Package".

CTX_DDL.CREATE_SECTION_GROUP in Chapter 7, "CTX_DDL Package".

ALTER INDEX

CATSEARCH
DROP INDEX

**Note:** This section describes the **DROP INDEX** statement as it pertains to dropping a Text domain index.

For a complete description of the **DROP INDEX** statement, see *Oracle Database SQL Language Reference*.

**Purpose**

Use **DROP INDEX** to drop a specified Text index.

**Syntax**

```
DROP INDEX [schema.]index [force];
```

**[force]**

Optionally force the index to be dropped. Use **force** option when Oracle Text cannot determine the state of the index, such as when an indexing operation crashes.

Oracle recommends against using this option by default. Use it as a last resort when a regular call to **DROP INDEX** fails.

**Example**

The following example drops an index named **doc_index** in the current user's database schema.

```
DROP INDEX doc_index;
```

**Related Topics**

- **ALTER INDEX**
- **CREATE INDEX**
MATCHES

Use the MATCHES operator to find all rows in a query table that match a given document. The document must be a plain text, HTML, or XML document.

The MATCHES operator also supports database links. You can identify a remote table or materialized view by appending @dblink to the end of its name. The dblink must be a complete or partial name for a database link to the database containing the remote table or materialized view. (Querying of remote views is not supported.)

This operator requires a CTXRULE index on your set of queries.

When the SVM_CLASSIFIER classifier type is used, MATCHES returns a score in the range 0 to 100; a higher number indicates a greater confidence in the match. Use the label parameter and MATCH_SCORE to obtain this number. Then use the matching score to apply a category-specific threshold to a particular category.

If SVM_CLASSIFIER is not used, then this operator returns either 100 (the document matches the criteria) or 0 (the document does not match).

Limitation

If the optimizer chooses to use the functional query invocation with a MATCHES query, your query will fail.

Syntax

MATCHES(
[schema.]column,
document VARCHAR2 or CLOB
[,label INTEGER])
RETURN NUMBER;

column
Specify the column containing the indexed query set.

document
Specify the document to be classified. The document can be plain-text, HTML, or XML. Binary formats are not supported.

label
Optionally specify the label that identifies the score generated by the MATCHES operator. Use this label with MATCH_SCORE.

Matches Example

The following example creates a table querytable, and populates it with classification names and associated rules. It then creates a CTXRULE index.

The example enters the MATCHES query with a document string to be classified. The SELECT statement returns all rows (queries) that are satisfied by the document:

create table querytable (classification varchar2(64), text varchar2(4000));
insert into querytable values ('common names', 'smith OR jones OR brown');
insert into querytable values ('countries', 'United States OR Great Britain OR France');
insert into querytable values ('Oracle DB', 'oracle NEAR database');
create index query_rule on querytable(text) indextype is ctxsys.ctxrule;
SELECT classification FROM querytable WHERE MATCHES(text, 'Smith is a common name in the United States') > 0;

CLASSIFICATION
-----------------------------------------------
common names
countries

Related Topics

MATCH_SCORE on page 1-61
Syntax for CTXRULE Index Type on page 1-55
CTX_CLS.TRAIN on page 6-2

The Oracle Text Application Developer’s Guide contains extended examples of simple and supervised classification, which make use of the MATCHES operator.
MATCH_SCORE

Use the MATCH_SCORE operator in a statement to return scores produced by a MATCHES query.

The MATCH_SCORE operator also supports database links. You can identify a remote table or materialized view by appending @dblink to the end of its name. The dblink must be a complete or partial name for a database link to the database containing the remote table or materialized view. (Querying of remote views is not supported.)

When the SVM_CLASSIFIER classifier type is used, this operator returns a score in the range 0 to 100. Then use the matching score to apply a category-specific threshold to a particular category.

If SVM_CLASSIFIER is not used, then this operator returns either 100 (the document matches the criteria) or 0 (the document does not match).

Syntax

MATCH_SCORE(label NUMBER)

label
Specify a number to identify the score produced by the query. Use this number to identify the MATCHES clause that returns this score.

Example

To get the matching score, use

select cat_id, match_score(1) from training_result where matches(profile, text,1)>0;

Related Topics

MATCHES on page 1-59
Use the `SCORE` operator in a `SELECT` statement to return the score values produced by a `CONTAINS` query. The `SCORE` operator can be used in a `SELECT`, `ORDER BY`, or `GROUP BY` clause.

The `SCORE` operator also supports database links. You can identify a remote table or materialized view by appending `@dblink` to the end of its name. The `dblink` must be a complete or partial name for a database link to the database containing the remote table or materialized view. (Querying of remote views is not supported.)

**Syntax**

```
SCORE(label NUMBER)
```

**label**

Specify a number to identify the score produced by the query. Use this number to identify the `CONTAINS` clause which returns this score.

**Example**

**Single CONTAINS**

When the `SCORE` operator is called (for example, in a `SELECT` clause), the `CONTAINS` clause must reference the score label value as in the following example:

```
SELECT SCORE(1), title from newsindex
WHERE CONTAINS(text, 'oracle', 1) > 0 ORDER BY SCORE(1) DESC;
```

**Multiple CONTAINS**

Assume that a news database stores and indexes the title and body of news articles separately. The following query returns all the documents that include the words `Oracle` in their title and `java` in their body. The articles are sorted by the scores for the first `CONTAINS (Oracle)` and then by the scores for the second `CONTAINS (java)`.

```
SELECT title, body, SCORE(10), SCORE(20)
FROM news
WHERE CONTAINS (news.title, 'Oracle', 10) > 0 OR
CONTAINS (news.body, 'java', 20) > 0
ORDER BY SCORE(10), SCORE(20);
```

**Related Topics**

- `CONTAINS`
- Appendix F, "The Oracle Text Scoring Algorithm"
This chapter describes the indexing elements that you can use to create an Oracle Text index.

The following topics are discussed in this chapter:

- Overview
- Datastore Types
- Filter Types
- Lexer Types
- Wordlist Type
- Storage Types
- Section Group Types
- Classifier Types
- Cluster Types
- Stoplists
- System-Defined Preferences
- System Parameters

Overview

When you use CREATE INDEX to create an index or ALTER INDEX to manage an index, you can optionally specify indexing preferences, stoplists, and section groups in the parameter string. Specifying a preference, stoplist, or section group answers one of the following questions about the way Oracle Text indexes text:

<table>
<thead>
<tr>
<th>Preference Class</th>
<th>Answers the Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datastore</td>
<td>How are your documents stored?</td>
</tr>
<tr>
<td>Filter</td>
<td>How can the documents be converted to plain text?</td>
</tr>
<tr>
<td>Lexer</td>
<td>What language is being indexed?</td>
</tr>
<tr>
<td>Wordlist</td>
<td>How should stem and fuzzy queries be expanded?</td>
</tr>
<tr>
<td>Storage</td>
<td>How should the index tables be stored?</td>
</tr>
<tr>
<td>Stop List</td>
<td>What words or themes are not to be indexed?</td>
</tr>
</tbody>
</table>
This chapter describes how to set each preference. Enable an option by creating a preference with one of the types described in this chapter.

For example, to specify that your documents are stored in external files, you can create a datastore preference called mydatastore using the FILE_DATASTORE type. Specify `mydatastore` as the datastore preference in the parameter clause of `CREATE INDEX`.

### Creating Preferences

To create a datastore, lexer, filter, classifier, wordlist, or storage preference, use the CTX_DDL.CREATE_PREFERENCE procedure and specify one of the types described in this chapter. For some types, you can also set attributes with the CTX_DDL.SET_ATTRIBUTE procedure.

An indexing type names a class of indexing objects that you can use to create an index preference. A type, therefore, is an abstract ID, while a preference is an entity that corresponds to a type. Many system-defined preferences have the same name as types (for example, BASIC_LEXER), but exact correspondence is not guaranteed. For example, the DEFAULT_DATASTORE preference uses the DIRECT_DATASTORE type, and there is no system preference corresponding to the CHARSET_FILTER type. Be careful in assuming the existence or nature of either indexing types or system preferences.

You specify indexing preferences with `CREATE INDEX` and `ALTER INDEX`. Indexing preferences determine how your index is created. For example, lexer preferences indicate the language of the text to be indexed. You can create and specify your own user-defined preferences, or you can utilize system-defined preferences.

To create a stoplist, use CTX_DDL.CREATE_STOPLIST. Add stopwords to a stoplist with CTX_DDL.ADD_STOPWORD.

To create section groups, use CTX_DDL.CREATE_SECTION_GROUP and specify a section group type. Add sections to section groups with CTX_DDL.ADD_ZONE_SECTION or CTX_DDL.ADD_FIELD_SECTION.

### Datastore Types

Use the datastore types to specify how your text is stored. To create a datastore preference, you must use one of the datastore types described in Table 2–1.

<table>
<thead>
<tr>
<th>Datastore Type</th>
<th>Use When</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIRECT_DATASTORE</td>
<td>Data is stored internally in the text column. Each row is indexed as a single document.</td>
</tr>
<tr>
<td>MULTI_COLUMN_DATASTORE</td>
<td>Data is stored in a text table in more than one column. Columns are concatenated to create a virtual document, one for each row.</td>
</tr>
</tbody>
</table>
**DIRECT_DATASTORE**

Use the DIRECT_DATASTORE type for text stored directly in the text column, one document for each row. DIRECT_DATASTORE has no attributes.

The following column types are supported: CHAR, VARCHAR, VARCHAR2, BLOB, CLOB, BFILE, XMLType, and URIType.

**Note:** If your column is a BFILE, then the index owner must have read permission on all directories used by the BFILES.

**DIRECT_DATASTORE CLOB Example**

The following example creates a table with a CLOB column to store text data. It then populates two rows with text data and indexes the table using the system-defined preference CTXSYS.DEFAULT_DATASTORE.

```
create table mytable(id number primary key, docs clob);
insert into mytable values(111555,'this text will be indexed');
insert into mytable values(111556,'this is a direct_datastore example');
commit;
create index myindex on mytable(docs)
    indextype is ctxsys.context
    parameters ('DATASTORE CTXSYS.DEFAULT_DATASTORE');
```

**MULTI_COLUMN_DATASTORE**

Use the MULTI_COLUMN_DATASTORE datastore when your text is stored in more than one column. During indexing, the system concatenates the text columns, tags the column text, and indexes the text as a single document. The XML-like tagging is optional. You can also set the system to filter and concatenate binary columns.

MULTI_COLUMN_DATASTORE has the attributes shown in Table 2–2.
Table 2–2  MULTI_COLUMN_DATASTORE Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>columns</td>
<td>Specify a comma-delimited list of columns to be concatenated during indexing. You can also specify any expression allowable for the SELECT statement column list for the base table. This includes expressions, PL/SQL functions, column aliases, and so on. NUMBER and DATE column types are supported. They are converted to text before indexing using the default format mask. The TO_CHAR function can be used in the column list for formatting. RAW and BLOB columns are directly concatenated as binary data. LONG, LONG RAW, NCHAR, and NCLOB, nested table columns and collections are not supported. The column list is limited to 500 bytes.</td>
</tr>
<tr>
<td>filter</td>
<td>Specify a comma-delimited list of Y/N flags. Each flag corresponds to a column in the COLUMNS list and denotes whether to filter the column using the AUTO_FILTER. Specify one of the following allowable values: Y: Column is to be filtered with AUTO_FILTER N or no value: Column is not be filtered (Default)</td>
</tr>
<tr>
<td>delimiter</td>
<td>Specify the delimiter that separates column text as follows: COLUMN_NAME_TAG: Column text is set off by XML-like open and close tags (default behavior). NEWLINE: Column text is separated with a newline.</td>
</tr>
</tbody>
</table>

Indexing and DML

To index, you must create a dummy column to specify in the CREATE INDEX statement. This column’s contents are not made part of the virtual document, unless its name is specified in the columns attribute.

The index is synchronized only when the dummy column is updated. You can create triggers to propagate changes if needed.

MULTI_COLUMN_DATASTORE Restriction

You cannot create a multi-column datastore with XMLType columns. MULTI_COLUMN_DATASTORE does not support XMLType. You can create a CONTEXT index with an XMLType column, as described in Chapter 1, “Oracle Text SQL Statements and Operators”.

MULTI_COLUMN_DATASTORE Example

The following example creates a multi-column datastore preference called my_multi with three text columns:

```
begin
  ctx_ddl.create_preference('my_multi', 'MULTI_COLUMN_DATASTORE');
  ctx_ddl.set_attribute('my_multi', 'columns', 'column1, column2, column3');
end;
```

MULTI_COLUMN_DATASTORE Filter Example

The following example creates a multi-column datastore preference and denotes that the bar column is to be filtered with the AUTO_FILTER.

```
ctx_ddl.create_preference('MY_MULTI', 'MULTI_COLUMN_DATASTORE');
```
The multi-column datastore fetches the content of the foo and bar columns, filters bar, then composes the compound document as:

```xml
<FOO>
foo contents
</FOO>
<BAR>
bar filtered contents (probably originally HTML)
</BAR>
```

The N's need not be specified, and there need not be a flag for every column. Only the Y's need to be specified, with commas to denote which column they apply to. For instance:

```sql
ctx_ddl.create_preference('MY_MULTI','MULTI_COLUMN_DATASTORE');
ctx_ddl.set_attribute('MY_MULTI', 'COLUMNS','foo,bar,zoo,jar');
ctx_ddl.set_attribute('MY_MULTI','FILTER',',,Y');
```

This filters only the column zoo.

**Tagging Behavior**

During indexing, the system creates a virtual document for each row. The virtual document is composed of the contents of the columns concatenated in the listing order with column name tags automatically added. For example:

```sql
create table mc(id number primary key, name varchar2(10), address varchar2(80));
insert into mc values(1, 'John Smith', '123 Main Street');
exec ctx_ddl.create_preference('mymds', 'MULTI_COLUMN_DATASTORE');
exec ctx_ddl.set_attribute('mymds', 'columns', 'name, address');
```

This produces the following virtual text for indexing:

```xml
<NAME>
John Smith
</NAME>
<ADDRESS>
123 Main Street
</ADDRESS>
```

The system indexes the text between the tags, ignoring the tags themselves.

**Indexing Columns as Sections**

To index these tags as sections, you can optionally create field sections with the BASIC_SECTION_GROUP.

---

**Note:** No section group is created when you use the MULTI_COLUMN_DATASTORE. To create sections for these tags, you must create a section group.

---

When you use expressions or functions, the tag is composed of the first 30 characters of the expression unless a column alias is used.

For example, if your expression is as follows:
exec ctx_ddl.set_attribute('mymds', 'columns', '4 + 17');

then it produces the following virtual text:

<4 + 17>
21
</4 + 17>

If your expression is as follows:

exec ctx_ddl.set_attribute('mymds', 'columns', '4 + 17 col1');

then it produces the following virtual text:

<col1>
21
<col1>

The tags are in uppercase unless the column name or column alias is in lowercase and surrounded by double quotes. For example:

exec ctx_ddl.set_attribute('mymds', 'COLUMNS', 'foo');

produces the following virtual text:

<F00>
content of foo
</F00>

For lowercase tags, use the following:

exec ctx_ddl.set_attribute('mymds', 'COLUMNS', 'foo "foo"');

This expression produces:

<f00>
content of foo
</f00>

**DETAIL_DATASTORE**

Use the **DETAIL_DATASTORE** type for text stored directly in the database in detail tables, with the indexed text column located in the master table.

**DETAIL_DATASTORE** has the following attributes:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>binary</td>
<td>Specify <strong>TRUE</strong> for Oracle Text to add <em>no</em> newline character after each detail row. Specify <strong>FALSE</strong> for Oracle Text to add a newline character ('\n') after each detail row automatically.</td>
</tr>
<tr>
<td>detail_table</td>
<td>Specify the name of the detail table (*OWNER.*TABLE if necessary).</td>
</tr>
<tr>
<td>detail_key</td>
<td>Specify the name of the detail table foreign key column(s).</td>
</tr>
<tr>
<td>detail_lineno</td>
<td>Specify the name of the detail table sequence column.</td>
</tr>
<tr>
<td>detail_text</td>
<td>Specify the name of the detail table text column.</td>
</tr>
</tbody>
</table>
Synchronizing Master/Detail Indexes
Changes to the detail table do not trigger re-indexing when you synchronize the index. Only changes to the indexed column in the master table triggers a re-index when you synchronize the index.

You can create triggers on the detail table to propagate changes to the indexed column in the master table row.

Example Master/Detail Tables
This example illustrates how master and detail tables are related to each other.

Master Table Example Master tables define the documents in a master/detail relationship. Assign an identifying number to each document. The following table is an example master table, called my_master:

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Column Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>article_id</td>
<td>NUMBER</td>
<td>Document ID, unique for each document (Primary Key)</td>
</tr>
<tr>
<td>author</td>
<td>VARCHAR2(30)</td>
<td>Author of document</td>
</tr>
<tr>
<td>title</td>
<td>VARCHAR2(50)</td>
<td>Title of document</td>
</tr>
<tr>
<td>body</td>
<td>CHAR(1)</td>
<td>Dummy column to specify in CREATE INDEX</td>
</tr>
</tbody>
</table>

Note: Your master table must include a primary key column when you use the DETAIL_DATASTORE type.

Detail Table Example Detail tables contain the text for a document, whose content is usually stored across a number of rows. The following detail table my_detail is related to the master table my_master with the article_id column. This column identifies the master document to which each detail row (sub-document) belongs.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Column Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>article_id</td>
<td>NUMBER</td>
<td>Document ID that relates to master table</td>
</tr>
<tr>
<td>seq</td>
<td>NUMBER</td>
<td>Sequence of document in the master document defined by article_id</td>
</tr>
<tr>
<td>text</td>
<td>VARCHAR2</td>
<td>Document text</td>
</tr>
</tbody>
</table>

Detail Table Example Attributes In this example, the DETAIL_DATASTORE attributes have the following values:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>binary</td>
<td>TRUE</td>
</tr>
<tr>
<td>detail_table</td>
<td>my_detail</td>
</tr>
<tr>
<td>detail_key</td>
<td>article_id</td>
</tr>
<tr>
<td>detail_lineno</td>
<td>seq</td>
</tr>
<tr>
<td>detail_text</td>
<td>text</td>
</tr>
</tbody>
</table>
Use CTX_DDL.CREATE_PREFERENCE to create a preference with DETAIL_DATASTORE. Use CTX_DDL.SET_ATTRIBUTE to set the attributes for this preference as described earlier. The following example shows how this is done:

```
begin
  ctx_ddl.create_preference('my_detail_pref', 'DETAIL_DATASTORE');
  ctx_ddl.set_attribute('my_detail_pref', 'binary', 'true');
  ctx_ddl.set_attribute('my_detail_pref', 'detail_table', 'my_detail');
  ctx_ddl.set_attribute('my_detail_pref', 'detail_key', 'article_id');
  ctx_ddl.set_attribute('my_detail_pref', 'detail_lineno', 'seq');
  ctx_ddl.set_attribute('my_detail_pref', 'detail_text', 'text');
end;
```

**Master/Detail Index Example** To index the document defined in this master/detail relationship, specify a column in the master table with CREATE_INDEX. The column you specify must be one of the allowable types.

This example uses the body column, whose function is to enable the creation of the master/detail index and to improve readability of the code. The my_detail_pref preference is set to DETAIL_DATASTORE with the required attributes:

```
CREATE INDEX myindex on my_master(body) indextype is ctxsys.context
parameters('datastore my_detail_pref');
```

In this example, you can also specify the title or author column to create the index. However, if you do so, changes to these columns will trigger a re-index operation.

**FILE_DATASTORE**

The FILE_DATASTORE type is used for text stored in files accessed through the local file system.

---

**Note:** FILE_DATASTORE may not work with certain types of remote mounted file systems.

---

FILE_DATASTORE has the following attribute(s):

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>path</td>
<td>path1: path2: pathn</td>
</tr>
<tr>
<td>filename_charset</td>
<td>name</td>
</tr>
</tbody>
</table>

**path**

Specify the full directory path name of the files stored externally in a file system. When you specify the full directory path as such, you need only include file names in your text column.

You can specify multiple paths for path, with each path separated by a colon (:) on UNIX and semicolon (;) on Windows. File names are stored in the text column in the text table.

If you do not specify a path for external files with this attribute, then Oracle Text requires that the path be included in the file names stored in the text column.
filename_charset
Specify a valid Oracle character set name (maximum length 30 characters) to be used by the file datastore for converting filenames. In general, the Oracle database can use a different character set than the operating system. This can lead to problems in finding files (which may raise DRG-11513 errors) when the indexed column contains characters that are not convertible to the operating system character set. By default, the file datastore will convert the file name to WE8ISO8859p1 for ASCII platforms or WE8EBCDIC1047 for EBCDIC platforms.

However, this may not be sufficient for applications with multibyte character sets for both the database and the operating system, since neither WE8ISO8859p1 nor WE8EBCDIC1047 supports multibyte characters. The attribute filename_charset rectifies this problem. If specified, then the datastore will convert from the database character set to the specified character set rather than to ISO8859 or EBCDIC.

If filename_charset is the same as the database character set, then the filename is used as is. If filename_charset is not a valid character set, then the error "DRG-10763: value %s is not a valid character set" is be raised.

PATH Attribute Limitations
The PATH attribute has the following limitations:

- If you specify a PATH attribute, then you can only use a simple filename in the indexed column. You cannot combine the PATH attribute with a path as part of the filename. If the files exist in multiple folders or directories, you must leave the PATH attribute unset, and include the full file name, with PATH, in the indexed column.

- On Windows systems, the files must be located on a local drive. They cannot be on a remote drive, whether the remote drive is mapped to a local drive letter.

FILE_DATASTORE and Security
File and URL datastores enable access to files on the actual database disk. This may be undesirable when security is an issue since any user can browse the file system that is accessible to the Oracle user. The FILE_ACCESS_ROLE system parameter can be used to set the name of a database role that is authorized to create an index using FILE or URL datastores. If set, any user attempting to create an index using FILE or URL datastores must have this role, or the index creation will fail. Only SYS can set FILE_ACCESS_ROLE, and an error will be raised if any other user tries to modify it. If FILE_ACCESS_ROLE is left at the default of NULL, access is disallowed. Thus, by default, users are not able to create indexes that use the file or URL datastores. Users can, if desired, set FILE_ACCESS_ROLE to PUBLIC if they want to preserve the behavior from earlier releases.

For example, the following statement sets the name of the database role:

cctxadm.set_parameter('FILE_ACCESS_ROLE','TOPCAT');

where TOPCAT is the role that is authorized to create an index on a file or URL datastore. The CREATE INDEX operation will fail when a user that does not have an authorized role tries to create an index on a file or URL datastore. For example:

CREATE INDEX myindex ON mydocument(TEXT) INDEXTYPE IS ctxsys.context
PARAMETERS('DATASTORE ctxsys.file_datastore')

In this case, if the user does not have the role TOPCAT, then index creation will fail and return an error. For users who have the TOPCAT role, the index creation will proceed normally.
The authorized role name is checked any time the datastore is accessed. This includes index creation, index sync, and calls to document services, such as `CTX_DOC.HIGHLIGHT`.

**FILE_DATASTORE Example**
This example creates a file datastore preference called `COMMON_DIR` that has a path of `/mydocs`:

```sql
begin
  ctx_ddl.create_preference('COMMON_DIR','FILE_DATASTORE');
  ctx_ddl.set_attribute('COMMON_DIR','PATH','/mydocs');
end;
```

When you populate the table `mytable`, you need only insert filenames. The `path` attribute tells the system where to look during the indexing operation.

```sql
create table mytable(id number primary key, docs varchar2(2000));
insert into mytable values(111555,'first.txt');
insert into mytable values(111556,'second.txt');
commit;
```

Create the index as follows:

```sql
create index myindex on mytable(docs)
  indextype is ctxsys.context
  parameters ('datastore COMMON_DIR');
```

**URL_DATASTORE**
Use the `URL_DATASTORE` type for text stored:

- In files on the World Wide Web (accessed through HTTP or FTP)
- In files in the local file system (accessed through the file protocol)

Store each URL in a single text field.

**URL Syntax**
The syntax of a URL you store in a text field is as follows (with brackets indicating optional parameters):

```
[URL:]<access_scheme>://<host_name>[:<port_number>]/<url_path>
```

The `access_scheme` string can be either `ftp`, `http`, or `file`. For example:

```sql
http://mymachine.us.oracle.com/home.html
```

**Note:** The `login:password@` syntax within the URL is supported only for the `ftp` access scheme.

Because this syntax is partially compliant with the RFC 1738 specification, the following restriction holds for the URL syntax: The URL must contain only printable ASCII characters. Non-printable ASCII characters and multibyte characters must be escaped with the `%xx` notation, where `xx` is the hexadecimal representation of the special character.
**URL_DATASTORE Attributes**

URL_DATASTORE has the following attributes:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>timeout</td>
<td>The value of this attribute is ignored. This is provided for backward compatibility.</td>
</tr>
<tr>
<td>maxthreads</td>
<td>The value of this attribute is ignored. URL_DATASTORE is single-threaded. This is provided for backward compatibility.</td>
</tr>
<tr>
<td>urlsize</td>
<td>The value of this attribute is ignored. This is provided for backward compatibility.</td>
</tr>
<tr>
<td>maxurls</td>
<td>The value of this attribute is ignored. This is provided for backward compatibility.</td>
</tr>
<tr>
<td>maxdocsize</td>
<td>The value of this attribute is ignored. This is provided for backward compatibility.</td>
</tr>
<tr>
<td>http_proxy</td>
<td>Specify the fully qualified name of the host machine that serves as the HTTP proxy (gateway) for the machine on which Oracle Text is installed. You can optionally specify port number with a colon in the form hostname:port. You must set this attribute if the machine is in an intranet that requires authentication through a proxy server to access Web files located outside the firewall.</td>
</tr>
<tr>
<td>ftp_proxy</td>
<td>Specify the fully qualified name of the host machine that serves as the FTP proxy (gateway) for the server on which Oracle Text is installed. You can optionally specify a port number with a colon in the form hostname:port. This attribute must be set if the machine is in an intranet that requires authentication through a proxy server to access Web files located outside the firewall.</td>
</tr>
<tr>
<td>no_proxy</td>
<td>Specify the domain for no proxy server. Use a comma separated string of up to 16 domain names.</td>
</tr>
</tbody>
</table>
**no_proxy**

Specify a string of domains (up to sixteen, separated by commas) that are found in most, if not all, of the machines in your intranet. When one of the domains is encountered in a host name, no request is sent to the server(s) specified for `ftp_proxy` and `http_proxy`. Instead, the request is processed directly by the host machine identified in the URL.

For example, if the string `us.example.com, uk.example.com` is entered for `no_proxy`, any URL requests to machines that contain either of these domains in their host names are not processed by your proxy server(s).

**URL_DATASTORE and Security**

For a discussion of how to control file access security for file and URL datastores, refer to “FILE_DATASTORE and Security” on page 2-9.

**URL_DATASTORE Example**

This example creates a `URL_DATASTORE` preference called `URL_PREF` for which the `http_proxy`, `no_proxy`, and `timeout` attributes are set. The defaults are used for the attributes that are not set.

```plaintext
begin
    ctx_ddl.create_preference('URL_PREF','URL_DATASTORE');
    ctx_ddl.set_attribute('URL_PREF','HTTP_PROXY','www-proxy.us.oracle.com');
    ctx_ddl.set_attribute('URL_PREF','NO_PROXY','us.oracle.com');
    ctx_ddl.set_attribute('URL_PREF','Timeout','300');
    end;
```

Create the table and insert values into it:

```plaintext
create table urls(id number primary key, docs varchar2(2000));
insert into urls values(111555,'http://context.us.oracle.com');
insert into urls values(111556,'http://www.sun.com');
commit;
```

To create the index, specify `URL_PREF` as the datastore:

```plaintext
create index datastores_text on urls ( docs )
    indextype is ctxsys.context
    parameters ( 'Datastore URL_PREF' );
```

**USER_DATASTORE**

Use the `USER_DATASTORE` type to define stored procedures that synthesize documents during indexing. For example, a user procedure might synthesize author, date, and text columns into one document to have the author and date information be part of the indexed text.

`USER_DATASTORE` has the following attributes:

**Table 2-6   USER_DATASTORE Attributes**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>procedure</td>
<td>Specify the procedure that synthesizes the document to be indexed.</td>
</tr>
<tr>
<td></td>
<td>This procedure can be owned by any user and must be executable by the index owner.</td>
</tr>
</tbody>
</table>
**Datastore Types**

### Oracle Text Indexing Elements

**procedure**

Specify the name of the procedure that synthesizes the document to be indexed. This specification must be in the form `PROCEDURENAME` or `PACKAGENAME.PROCEDURENAME`. You can also specify the schema owner name.

The procedure you specify must have two arguments defined as follows:

```
procedure (r IN ROWID, c IN OUT NOCOPY output_type)
```

The first argument `r` must be of type `ROWID`. The second argument `c` must be of type `output_type`. `NOCOPY` is a compiler hint that instructs Oracle Text to pass parameter `c` by reference if possible.

**Note:** The procedure name and its arguments can be named anything. The arguments `r` and `c` are used in this example for simplicity.

The stored procedure is called once for each row indexed. Given the rowid of the current row, `procedure` must write the text of the document into its second argument, whose type you specify with `output_type`.

**Constraints**

The following constraints apply to `procedure`:

- It can be owned by any user, but the user must have database permissions to execute `procedure` correctly
- It must be executable by the index owner
- It must not enter DDL or transaction control statements, like `COMMIT`

**Editing Procedure after Indexing**

When you change or edit the stored procedure, indexes based on it will not be notified, so you must manually re-create such indexes. So if the stored procedure makes use of other columns, and those column values change, the row will not be re-indexed. The row is re-indexed only when the indexed column changes.

**output_type**

Specify the datatype of the second argument to `procedure`. You can use either `CLOB`, `BLOB`, `CLOB_LOC`, `BLOB_LOC`, or `VARCHAR2`. The default is `CLOB`.

When you specify `CLOB_LOC`, `BLOB_LOC`, you indicate that no temporary `CLOB` or `BLOB` is needed, because your procedure copies a locator to the `IN/OUT` second parameter.

**USER_DATASTORE with CLOB Example**

Consider a table in which the author, title, and text fields are separate, as in the `articles` table defined as follows:

```
create table articles(
   id      number,
```

---

**Table 2-6 (Cont.) USER_DATASTORE Attributes**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>output_type</td>
<td>Specify the data type of the second argument to procedure. Valid values are <code>CLOB</code>, <code>BLOB</code>, <code>CLOB_LOC</code>, <code>BLOB_LOC</code>, or <code>VARCHAR2</code>. The default is <code>CLOB</code>. When you specify <code>CLOB_LOC</code>, <code>BLOB_LOC</code>, you indicate that no temporary <code>CLOB</code> or <code>BLOB</code> is needed, because your procedure copies a locator to the <code>IN/OUT</code> second parameter.</td>
</tr>
</tbody>
</table>

---
author  varchar2(80),
title   varchar2(120),
text    clob );

The author and title fields are to be part of the indexed document text. Assume user appowner writes a stored procedure with the user datastore interface that synthesizes a document from the text, author, and title fields:

create procedure myproc(rid in rowid, tlob in out clob nocopy) is
begin
  for c1 in (select author, title, text from articles
    where rowid = rid)
  loop
    dbms_lob.writeappend(tlob, length(c1.title), c1.title);
    dbms_lob.writeappend(tlob, length(c1.author), c1.author);
    dbms_lob.writeappend(tlob, length(c1.text), c1.text);
  end loop;
end;

This procedure takes in a rowid and a temporary CLOB locator, and concatenates all the article's columns into the temporary CLOB. The for loop executes only once.

The user appowner creates the preference as follows:

begin
  ctx_ddl.create_preference('myud', 'user_datastore');
  ctx_ddl.set_attribute('myud', 'procedure', 'myproc');
  ctx_ddl.set_attribute('myud', 'output_type', 'CLOB');
end;

When appowner creates the index on articles(text) using this preference, the indexing operation sees author and title in the document text.

**USER_DATASTORE with BLOB_LOC Example**

The following procedure might be used with `OUTPUT_TYPE BLOB_LOC`:

procedure myds(rid in rowid, dataout in out nocopy blob) is
  l_dtype varchar2(10);
  l_pk number;
begin
  select dtype, pk into l_dtype, l_pk from mytable where rowid = rid;
  if (l_dtype = 'MOVIE') then
    select movie_data into dataout from movietab where fk = l_pk;
  elsif (l_dtype = 'SOUND') then
    select sound_data into dataout from soundtab where fk = l_pk;
  end if;
end;

The user appowner creates the preference as follows:

begin
  ctx_ddl.create_preference('myud', 'user_datastore');
  ctx_ddl.set_attribute('myud', 'procedure', 'myproc');
  ctx_ddl.set_attribute('myud', 'output_type', 'blob_loc');
end;

**NESTED_DATASTORE**

Use the nested datastore type to index documents stored as rows in a nested table.
When using the nested table datastore, you must index a dummy column, because the extensible indexing framework disallows indexing the nested table column. See the example.

DML on the nested table is not automatically propagated to the dummy column used for indexing. For DML on the nested table to be propagated to the dummy column, your application code or trigger must explicitly update the dummy column.

Filter defaults for the index are based on the type of the nested_text column.

During validation, Oracle Text checks that the type exists and that the attributes you specify for nested_lineno and nested_text exist in the nested table type. Oracle Text does not check that the named nested table column exists in the indexed table.

**NESTED_DATASTORE Example**

This section shows an example of using the `NESTED_DATASTORE` type to index documents stored as rows in a nested table.

**Create the Nested Table** The following code creates a nested table and a storage table `mytab` for the nested table:

```sql
create type nt_rec as object {
  lno number, -- line number
  ltxt varchar2(80) -- text of line
};

create type nt_tab as table of nt_rec;
create table mytab {
  id number primary key, -- primary key
  dummy char(1), -- dummy column for indexing
  doc nt_tab -- nested table
}
nested table doc store as myntab;
```

**Insert Values into Nested Table** The following code inserts values into the nested table for the parent row with ID equal to 1.

```sql
insert into mytab values (1, null, nt_tab());
```
Filter Types

Use the filter types to create preferences that determine how text is filtered for indexing. Filters enable word processor documents, formatted documents, plain text, HTML, and XML documents to be indexed.

For formatted documents, Oracle Text stores documents in their native format and uses filters to build interim plain text or HTML versions of the documents. Oracle Text indexes the words derived from the plain text or HTML version of the formatted document. The TMP_DIR environment variable sets the directory path for storing temporary files created by the filter.

To create a filter preference, you must use one of the following types:

Table 2–8 Filter Types

<table>
<thead>
<tr>
<th>Filter</th>
<th>When Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHARSET_FILTER</td>
<td>Character set converting filter.</td>
</tr>
<tr>
<td>AUTO_FILTER</td>
<td>Auto filter for filtering formatted documents.</td>
</tr>
<tr>
<td>NULL_FILTER</td>
<td>No filtering required. Use for indexing plain text, HTML, or XML documents.</td>
</tr>
</tbody>
</table>

Create Nested Table Preferences The following code sets the preferences and attributes for the NESTED_DATASTORE according to the definitions of the nested table type nt_tab and the parent table mytab:

begin
-- create nested datastore pref
ctx_ddl.create_preference('ntds','nested_datastore');

-- nest tab column in main table
ctx_ddl.set_attribute('ntds','nested_column', 'doc');

-- nested table type
ctx_ddl.set_attribute('ntds','nested_type', 'scott.nt_tab');

-- lineno column in nested table
ctx_ddl.set_attribute('ntds','nested_lineno','lno');

--text column in nested table
ctx_ddl.set_attribute('ntds','nested_text', 'ltxt');
end;

Create Index on Nested Table The following code creates the index using the nested table datastore:

create index myidx on mytab(dummy) -- index dummy column, not nest table
indextype is ctxsys.context parameters ('datastore ntds');

Query Nested Datastore The following select statement queries the index built from a nested table:

select * from mytab where contains(dummy, 'dog and mat')>0;
-- returns document 1, because it has dog in line 1 and mat in line 2.

Filter Types
Use the CHARSET_FILTER to convert documents from a non-database character set to the character set used by the database.

CHARSET_FILTER has the attribute described in Table 2–9.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>charset</td>
<td>Specify the Globalization Support name of source character set.</td>
</tr>
<tr>
<td></td>
<td>If you specify UTF16AUTO, then this filter automatically detects the if the character set is UTF16 big- or little-endian.</td>
</tr>
<tr>
<td></td>
<td>Specify JAAUTO for Japanese character set auto-detection. This filter automatically detects the custom character specification in JA16EUC or JA16SJIS and converts to the database character set. This filter is useful in Japanese when your data files have mixed character sets.</td>
</tr>
<tr>
<td></td>
<td>JAAUTO can only be specified on a database whose character set is JA16EUC, JA16SJIS, or UTF8.</td>
</tr>
<tr>
<td></td>
<td>Specify AUTO to have CHARSET_FILTER automatically detect and convert character sets that Oracle Database supports, as shown in Table 2–10.</td>
</tr>
</tbody>
</table>

When the charset column or attribute is set to AUTO, the CHARSET_FILTER automatically detects the document character set and converts the document from the detected character set to the database character set. CHARSET_FILTER can detect the supported character sets shown in Table 2–10.

| Character Set |/
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AL16UTF16</td>
</tr>
<tr>
<td>AL32UTF8</td>
</tr>
<tr>
<td>AR8ISO8859P6</td>
</tr>
<tr>
<td>AR8MSWIN1256</td>
</tr>
<tr>
<td>CL8ISO8859P5</td>
</tr>
<tr>
<td>CL8KOI8R</td>
</tr>
<tr>
<td>CL8MSWIN1251</td>
</tr>
<tr>
<td>EE8ISO8859P2</td>
</tr>
<tr>
<td>EE8MSWIN1250</td>
</tr>
<tr>
<td>EL8ISO8859P7</td>
</tr>
<tr>
<td>EL8MSWIN1253</td>
</tr>
</tbody>
</table>
UTF-16 Big- and Little-Endian Detection
If your character set is UTF-16, then you can specify UTF16AUTO to automatically detect big- or little-endian data. Oracle Text does so by examining the first two bytes of the document row.

If the first two bytes are 0xFE, 0xFF, the document is recognized as big-endian and the remainder of the document minus those two bytes is passed on for indexing.

If the first two bytes are 0xFF, 0xFE, the document is recognized as little-endian and the remainder of the document minus those two bytes is passed on for indexing.

If the first two bytes are anything else, the document is assumed to be big-endian and the whole document including the first two bytes is passed on for indexing.

Indexing Mixed-Character Set Columns
A mixed character set column is one that stores documents of different character sets. For example, a text table might store some documents in WE8ISO8859P1 and others in UTF8.

To index a table of documents in different character sets, you must create your base table with a character set column. In this column, specify the document character set on a per-row basis. To index the documents, Oracle Text converts the documents into the database character set.

Character set conversion works with the CHARSET_FILTER. When the charset column is NULL or not recognized, Oracle Text assumes the source character set is the one specified in the charset attribute.

---

Note: Character set conversion also works with the AUTO_FILTER when the document format column is set to TEXT.

---

Indexing Mixed-Character Set Example
For example, create the table with a charset column:

```sql
create table hdocs (    id number primary key,    fmt varchar2(10),    cset varchar2(20),    text varchar2(80) ) ;
```

Create a preference for this filter:

```sql
begin    cxt_ddl.create_preference('cs_filter', 'CHARSET_FILTER');    cxt_ddl.set_attribute('cs_filter', 'charset', 'UTF8'); end; /
```

Insert plain-text documents and name the character set:

```sql
insert into hdocs values(1, 'text', 'WE8ISO8859P1', '/docs/iso.txt'); insert into hdocs values (2, 'text', 'UTF8', '/docs/utf8.txt'); commit;
```
Create the index and name the charset column:

```sql
create index hdocsx on hdocs(text) indextype is ctxsys.context
parameters ('datastore ctxsys.file_datastore
  filter cs_filter
  format column fmt
  charset column cset');
```

**AUTO_FILTER**

The AUTO_FILTER is a universal filter that filters most document formats, including PDF and Microsoft Word documents. Use it for indexing both single-format and mixed-format columns. This filter automatically bypasses plain text, HTML, XHTML, SGML, and XML documents.

**See Also:** Appendix B, "Oracle Text Supported Document Formats", for a list of the formats supported by AUTO_FILTER, and to learn more about how to set up your environment.

**Note:** The AUTO_FILTER replaces the INSO_FILTER, which has been deprecated. While every effort has been made to ensure maximal backward compatibility between the two filters, so that applications using INSO_FILTER will continue to work without modification, some differences may arise. Users should therefore use AUTO_FILTER in their new programs and, when possible, replace instances of INSO_FILTER, and any system preferences or constants that make use of it, in older applications.

The AUTO_FILTER preference has the following attributes:

**Table 2–11 AUTO_FILTER Attributes**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>timeout</td>
<td>Specify the AUTO_FILTER timeout in seconds. Use a number between 0 and 42,949,672. Default is 120. Setting this value to 0 disables the feature. How this wait period is used depends on how you set timeout_type. This feature is disabled for rows for which the corresponding charset and format column cause the AUTO_FILTER to bypass the row, such as when format is marked TEXT. Use this feature to prevent the Oracle Text indexing operation from waiting indefinitely on a hanging filter operation.</td>
</tr>
<tr>
<td>timeout_type</td>
<td>Specify either HEURISTIC or FIXED. Default is HEURISTIC. Specify HEURISTIC for Oracle Text to check every TIMEOUT seconds if output from Outside In HTML Export has increased. The operation terminates for the document if output has not increased. An error is recorded in the CTX_USER_INDEX_ERRORS view and Oracle Text moves to the next document row to be indexed. Specify FIXED to terminate the Outside In HTML Export processing after TIMEOUT seconds regardless of whether filtering was progressing normally or just hanging. This value is useful when indexing throughput is more important than taking the time to successfully filter large documents.</td>
</tr>
</tbody>
</table>
Indexing Formatted Documents

To index a text column containing formatted documents such as Microsoft Word, use the AUTO_FILTER. This filter automatically detects the document format. Use the CTXSYS.AUTO_FILTER system-defined preference in the parameter clause as follows:

```sql
create index hdocsx on hdocs(text) indextype is ctxsys.context
parameters ('datastore ctxsys.file_datastore
filter ctxsys.auto_filter');
```

**Note:** The CTXSYS.AUTO_FILTER replaces CTXSYS.INSO_FILTER, which has been deprecated. Programs making use of CTXSYS.INSO_FILTER should still work. New programs should use CTXSYS.AUTO_FILTER.

Explicitly Bypassing Plain Text or HTML in Mixed Format Columns

A mixed-format column is a text column containing more than one document format, such as a column that contains Microsoft Word, PDF, plain text, and HTML documents.

The AUTO_FILTER can index mixed-format columns, automatically bypassing plain text, HTML, and XML documents. However, if you prefer not to depend on the built-in bypass mechanism, you can explicitly tag your rows as text and cause the AUTO_FILTER to ignore the row and not process the document in any way.

The format column in the base table enables you to specify the type of document contained in the text column. You can specify the following document types: TEXT, BINARY, and IGNORE. During indexing, the AUTO_FILTER ignores any document typed TEXT, assuming the charset column is not specified. (The difference between a document with a TEXT format column type and one with an IGNORE type is that the TEXT document is indexed, but ignored by the filter, while the IGNORE document is not indexed at all. Use IGNORE to overlook documents such as image files, or documents in a language that you do not want to index. IGNORE can be used with any filter type.)

To set up the AUTO_FILTER bypass mechanism, you must create a format column in your base table.

For example:

```sql
create table hdocs {
    id number primary key,
    fmt varchar2(10),
    text varchar2(80)
};
```

Assuming you are indexing mostly Word documents, you specify BINARY in the format column to filter the Word documents. Alternatively, to have the AUTO_FILTER ignore an HTML document, specify TEXT in the format column.

For example, the following statements add two documents to the text table, assigning one format as BINARY and the other TEXT:

```sql
insert into hdocs values(1, 'binary', '/docs/myword.doc');
```
insert in hdocs values (2, 'text', '/docs/index.html');
commit;

To create the index, use CREATE INDEX and specify the format column name in the parameter string:

create index hdocsx on hdocs(text) indextype is ctxsys.context
         parameters ('datastore ctxsys.file_datastore
                        filter ctxsys.auto_filter
                        format column fmt');

If you do not specify TEXT or BINARY for the format column, BINARY is used.

Note: You need not specify the format column in CREATE INDEX when using the AUTO_FILTER.

Character Set Conversion With AUTO_FILTER
The AUTO_FILTER converts documents to the database character set when the document format column is set to TEXT. In this case, the AUTO_FILTER looks at the charset column to determine the document character set.

If the charset column value is not an Oracle Text character set name, the document is passed through without any character set conversion.

Note: You need not specify the charset column when using the AUTO_FILTER.

If you do specify the charset column and do not specify the format column, the AUTO_FILTER works like the CHARSET_FILTER, except that in this case there is no Japanese character set auto-detection.

See Also:  "CHARSET_FILTER" on page 2-17.

NULL_FILTER
Use the NULL_FILTER type when plain text or HTML is to be indexed and no filtering needs to be performed. NULL_FILTER has no attributes.

Indexing HTML Documents
If your document set is entirely HTML, Oracle recommends that you use the NULL_FILTER in your filter preference.

For example, to index an HTML document set, specify the system-defined preferences for NULL_FILTER and HTML_SECTION_GROUP as follows:

create index myindex on docs(htmlfile) indextype is ctxsys.context
         parameters('filter ctxsys.null_filter
                        section group ctxsys.html_section_group');

See Also:  For more information on section groups and indexing HTML documents, see "Section Group Types" on page 2-77.

MAIL_FILTER
Use MAIL_FILTER to transform RFC-822, RFC-2045 messages into indexable text. The following limitations apply to the input:
Documents must be US-ASCII
Lines must not be longer than 1024 bytes
Documents must be syntactically valid with regard to RFC-822.

Behavior for invalid input is not defined. Some deviations may be robustly handled by the filter without error. Others may result in a fetch-time or filter-time error.

The MAIL_FILTER has the following attributes:

**Table 2–12 MAIL_FILTER Attributes**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INDEX_FIELDS</td>
<td>Specify a colon-separated list of fields to preserve in the output. These fields are transformed to tag markup. For example, if INDEX_FIELDS is set to &quot;FROM&quot;: From: Scott Tiger becomes: &lt;FROM&gt;Scott Tiger&lt;/FROM&gt; Only top-level fields are transformed in this way.</td>
</tr>
<tr>
<td>AUTO_FILTER_TIMEOUT</td>
<td>Specify a timeout value for the AUTO_FILTER filtering invoked by the mail filter. Default is 60. (Replaces the INSO_TIMEOUT attribute and is backward compatible with INSO_TIMEOUT.)</td>
</tr>
<tr>
<td>AUTO_FILTER_OUTPUT_FORMATTING</td>
<td>Specify either TRUE or FALSE. Default is TRUE. This attribute replaces the previous INSO_OUTPUT_FORMATTING attribute. However, it has no effect in the current release.</td>
</tr>
<tr>
<td>PART_FIELD_STYLE</td>
<td>Specify how fields occurring in lower-level parts and identified by the INDEX_FIELDS attribute should be transformed. The fields of the top-level message part identified by INDEX_FIELDS are always transformed to tag markup (see the previous description of INDEX_FIELDS); PART_FIELD_STYLE controls the transformation of subsequent parts; for example, attached e-mails. Possible values include IGNORE (the default), in which the part fields are not included for indexing; TAG, in which the part field names are transformed to tags, as occurs with top-level part fields; FIELD, in which the part field names are preserved as fields, not as tags; and TEXT, in which the part field names are eliminated and only the field content is preserved for indexing. See &quot;Mail Filter Example&quot; on page 2-24 for an example of how PART_FIELD_STYLE works.</td>
</tr>
</tbody>
</table>

**Filter Behavior**

This filter behaves in the following way for each document:

- Read and remove header fields
- Decode message body if needed, depending on Content-transfer-encoding field
- Take action depending on the Content-Type field value and the user-specified behavior specified in a mail filter configuration file. (See "About the Mail Filter Configuration File" on page 2-23.) The possible actions are:
  - produce the body in the output text (INCLUDE). If no character set is encountered in the INCLUDE parts in the Content-Type header field, then Oracle defaults to the value specified in the character set column in the base
table. Name your populated character set column in the parameter string of the `CREATE INDEX` command.

- **AUTO_FILTER** the body contents (`AUTO_FILTER` directive).
- remove the body contents from the output text (`IGNORE`)

If no behavior is specified for the type in the configuration file, then the defaults are as follows:

- `text/*`: produce body in the output text
- `application/*`: **AUTO_FILTER** the body contents
- `image/*, audio/*, video/*, model/*`: `IGNORE`

Multipart messages are parsed, and the mail filter applied recursively to each part. Each part is appended to the output.

All text produced will be charset-converted to the database character set, if needed.

### About the Mail Filter Configuration File

The `MAIL_FILTER` filter makes use of a mail filter configuration file, which contains directives specifying how a mail document should be filtered. The mail filter configuration file is a editable text file. Here you can override default behavior for each Content-Type. The configuration file also contains IANA-to-Oracle Globalization Support character set name mappings.

The location of the file must be in `ORACLE_HOME/ctx/config`. The name of the file to use is stored in the new system parameter `MAIL_FILTER_CONFIG_FILE`. On install, this is set to `dirmailfl.txt`, which has useful default contents.

Oracle recommends that you create your own mail filter configuration files to avoid overwrite by the installation of a new version or patch set. The mail filter configuration file should be in the database character set.

### Mail File Configuration File Structure

The file has two sections, **BEHAVIOR** and **CHARSETS**. Indicate the start of the behavior section as follows:

```
[behavior]
```

Each line following starts with a mime type, then whitespace, then behavior specification. The MIME type can be a full `TYPE/SUBTYPE` or just `TYPE`, which will apply to all subtypes of that type. `TYPE/SUBTYPE` specification overrides `TYPE` specification, which overrides default behavior. Behavior can be `INCLUDE`, `AUTO_FILTER`, or `IGNORE` (see "Filter Behavior" on page 2-22 for definitions). For instance:

```
application/zip IGNORE
application/msword AUTO_FILTER
model IGNORE
```

You cannot specify behavior for "multipart" or "message" types. If you do, such lines are ignored. Duplicate specification for a type replaces earlier specifications.

Comments can be included in the mail configuration file by starting lines with the `#` symbol.

The charset mapping section begins with

```
[charsets]
```
Lines consist of an IANA name, then whitespace, then an Oracle Globalization Support charset name, like:

```
US-ASCII     US7ASCII
ISO-8859-1   WB8ISO8859P1
```

This file is the only way the mail filter gets the mappings. There are no defaults. When you change the configuration file, the changes affect only the documents indexed after that point. You must flush the shared pool after changing the file.

**Mail_Filter Example**

Suppose there is an e-mail with the following form, in which other e-mails with different subject lines are attached to this e-mail:

```
To:  somebody@someplace
Subject:  mainheader
Content-Type:  multipart/mixed
.
.
Content-Type:  text/plain
X-Ref:  some_value
Subject:  subheader 1
.
.
Content-Type:  text/plain
X-Control:  blah blah blah
Subject:  subheader 2
.
.
Set INDEX_FIELDS to be "Subject" and, initially, PART_FIELD_STYLE to IGNORE.
```

```
CTX_DDL.CREATE_PREFERENCE('my_mail_filt', 'mail_filter');
CTX_DDL_SET_ATTRIBUTE('my_mail_filt', 'INDEX_FILES', 'subject');
CTX_DDL_SET_ATTRIBUTE('my_mail_filt', 'PART_FIELD_STYLE', 'ignore');
```

Now when the index is created, the file will be indexed as follows:

```
<SUBJECT>mainheader</SUBJECT>
```

If PART_FIELD_STYLE is instead set to TAG, this becomes:

```
<SUBJECT>mainheader</SUBJECT>
<SUBJECT>subheader1</SUBJECT>
<SUBJECT>subheader2</SUBJECT>
```

If PART_FIELD_STYLE is set to FIELD instead, this is the result:

```
<SUBJECT>mainheader</SUBJECT>
SUBJECT:subheader1
SUBJECT:subheader2
```

Finally, if PART_FIELD_STYLE is instead set to TEXT, then the result is:

```
<SUBJECT>mainheader</SUBJECT>
subheader1
subheader2
```

**USER_FILTER**

Use the USER_FILTER type to specify an external filter for filtering documents in a column. USER_FILTER has the following attribute:
command
Specify the executable for the single external filter that is used to filter all text stored in a column. If more than one document format is stored in the column, then the external filter specified for command must recognize and handle all such formats.

The executable that you specify must exist in the $ORACLE_HOME/ctx/bin directory on UNIX, and in the %ORACLE_HOME%/ctx/bin directory on Windows.

You must create your user-filter command with two parameters:

- The first parameter is the name of the input file to be read.
- The second parameter is the name of the output file to be written to.

If all the document formats are supported by AUTO_FILTER, then use AUTO_FILTER instead of USER_FILTER, unless additional tasks besides filtering are required for the documents.

Using USER_FILTER with Charset and Format Columns

USER_FILTER bypasses documents that do not need to be filtered. Its behavior is sensitive to the values of the format and charset columns. In addition, USER_FILTER performs character set conversion according to the charset column values.

Explicitly Bypassing Plain Text or HTML in Mixed Format Columns

A mixed-format column is a text column containing more than one document format, such as a column that contains Microsoft Word, PDF, plain text, and HTML documents.

The USER_FILTER executable can index mixed-format columns, automatically bypassing textual documents. However, if you prefer not to depend on the built-in bypass mechanism, you can explicitly tag your rows as text and cause the USER_FILTER executable to ignore the row and not process the document in any way.

The format column in the base table enables you to specify the type of document contained in the text column. You can specify the following document types: TEXT, BINARY, and IGNORE. During indexing, the USER_FILTER executable ignores any document typed TEXT, assuming the charset column is not specified. (The difference between a document with a TEXT format column type and one with an IGNORE type is that the TEXT document is indexed, but ignored by the filter, while the IGNORE document is not indexed at all. Use IGNORE to overlook documents such as image files, or documents in a language that you do not want to index. IGNORE can be used with any filter type.)

---

**Table 2–13  USER_FILTER Attribute**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>command</td>
<td>Specify the name of the filter executable.</td>
</tr>
</tbody>
</table>

---

**CAUTION:** The USER_FILTER type introduces the potential for security threats. A database user granted the CTXAPP role could potentially use USER_FILTER to load a malicious application. Therefore, the DBA must safeguard against any combination of input and output file parameters that would enable the named filter executable to compromise system security.
To set up the USER_FILTER bypass mechanism, you must create a format column in your base table. For example:

```sql
create table hdocs (
    id number primary key,
    fmt varchar2(10),
    text varchar2(80)
);
```

Assuming you are indexing mostly Word documents, you specify BINARY in the format column to filter the Word documents. Alternatively, to have the USER_FILTER executable ignore an HTML document, specify TEXT in the format column.

For example, the following statements add two documents to the text table, assigning one format as BINARY and the other TEXT:

```sql
insert into hdocs values(1, 'binary', '/docs/myword.doc');
insert into hdocs values(2, 'text', '/docs/index.html');
commit;
```

Assuming that this file is named upcase.pl, create the filter preference as follows:

```sql
ctx_ddl.create_preference
(  
    preference_name => 'USER_FILTER_PREF',
    object_name     => 'USER_FILTER'
);

ctx_ddl.set_attribute ('USER_FILTER_PREF', 'COMMAND', 'upcase.pl');
```

To create the index, use CREATE INDEX and specify the format column name in the parameter string:

```sql
create index hdocsx on hdocs(text) indextype is ctxsys.context
    parameters ('datastore ctxsys.file_datastore
    filter 'USER_FILTER_PREF'
    format column fmt');
```

If you do not specify TEXT or BINARY for the format column, BINARY is used.

**Character Set Conversion with USER_FILTER**

The USER_FILTER executable converts documents to the database character set when the document format column is set to TEXT. In this case, the USER_FILTER executable looks at the charset column to determine the document character set.

If the charset column value is not an Oracle Text character set name, the document is passed through without any character set conversion.

If you do specify the charset column and do not specify the format column, the USER_FILTER executable works like the CHARSET_FILTER, except that in this case, there is no Japanese character set auto-detection. See "CHARSET_FILTER" on page 2-17 for more information regarding CHARSET_FILTER.

**User Filter Example**

The following example shows a Perl script to be used as the user filter. This script converts the input text file specified in the first argument to uppercase and writes the output to the location specified in the second argument.

```perl
#!/usr/local/bin/perl
```
open(IN, $ARGV[0]);
open(OUT, '>'.$ARGV[1]);

while (<IN>)
{
    tr/a-z/A-Z/;
    print OUT;
}

close (IN);
close (OUT);

Assuming that this file is named upcase.pl, create the filter preference as follows:

begin
    ctx_ddl.create_preference
    {
        preference_name => 'USER_FILTER_PREF',
        object_name     => 'USER_FILTER'
    };
    ctx_ddl.set_attribute
    ('USER_FILTER_PREF','COMMAND','upcase.pl');
end;

Create the index in SQL*Plus as follows:

create index user_filter_idx on user_filter ( docs )
    indextype is ctxsys.context
    parameters ('FILTER USER_FILTER_PREF');

PROCEDURE_FILTER

Use the PROCEDURE_FILTER type to filter your documents with a stored procedure. The stored procedure is called each time a document needs to be filtered.

Table 2–14 lists the attributes for PROCEDURE_FILTER.

Table 2–14  PROCEDURE_FILTER Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Purpose</th>
<th>Allowable Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>procedure</td>
<td>Name of the filter stored procedure.</td>
<td>Any procedure. The procedure can be a PL/SQL stored procedure.</td>
</tr>
<tr>
<td>input_type</td>
<td>Type of input argument for stored procedure.</td>
<td>VARCHAR2, BLOB, CLOB, FILE</td>
</tr>
<tr>
<td>output_type</td>
<td>Type of output argument for stored procedure.</td>
<td>VARCHAR2, CLOB, FILE</td>
</tr>
<tr>
<td>rowid_parameter</td>
<td>Include rowid parameter?</td>
<td>TRUE/FALSE</td>
</tr>
<tr>
<td>format_parameter</td>
<td>Include format parameter?</td>
<td>TRUE/FALSE</td>
</tr>
<tr>
<td>charset_parameter</td>
<td>Include charset parameter?</td>
<td>TRUE/FALSE</td>
</tr>
</tbody>
</table>

**procedure**

Specify the name of the stored procedure to use for filtering. The procedure can be a PL/SQL stored procedure. The procedure can be a safe callout, or call a safe callout.
With the rowid_parameter, format_parameter, and charset_parameter set to FALSE, the procedure can have one of the following signatures:

- PROCEDURE(IN BLOB, IN OUT NOCOPY CLOB)
- PROCEDURE(IN CLOB, IN OUT NOCOPY CLOB)
- PROCEDURE(IN VARCHAR, IN OUT NOCOPY CLOB)
- PROCEDURE(IN BLOB, IN OUT NOCOPY VARCHAR2)
- PROCEDURE(IN CLOB, IN OUT NOCOPY VARCHAR2)
- PROCEDURE(IN VARCHAR2, IN OUT NOCOPY VARCHAR2)
- PROCEDURE(IN BLOB, IN VARCHAR2)
- PROCEDURE(IN CLOB, IN VARCHAR2)
- PROCEDURE(IN VARCHAR2, IN VARCHAR2)

The first argument is the content of the unfiltered row, output by the datastore. The second argument is for the procedure to pass back the filtered document text.

The procedure attribute is mandatory and has no default.

**input_type**
Specify the type of the input argument of the filter procedure. You can specify one of the following types:

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>procedure</td>
<td>Name of the filter stored procedure.</td>
</tr>
<tr>
<td>input_type</td>
<td>Type of input argument for stored procedure.</td>
</tr>
<tr>
<td>output_type</td>
<td>Type of output argument for stored procedure.</td>
</tr>
<tr>
<td>rowid_parameter</td>
<td>Include rowid parameter?</td>
</tr>
</tbody>
</table>

The input_type attribute is not mandatory. If not specified, then BLOB is the default.

**output_type**
Specify the type of output argument of the filter procedure. You can specify one of the following types:

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLOB</td>
<td>The output argument is IN OUT NOCOPY CLOB. Your procedure must write the filtered content to the CLOB passed in.</td>
</tr>
<tr>
<td>VARCHAR2</td>
<td>The output argument is IN OUT NOCOPY VARCHAR2. Your procedure must write the filtered content to the VARCHAR2 variable passed in.</td>
</tr>
<tr>
<td>FILE</td>
<td>The output argument must be VARCHAR2. On entering the filter procedure, the output argument is the name of a temporary file. The filter procedure must write the filtered contents to this named file. Using a FILE output type is useful only when the procedure is a safe callout, which can write to the file.</td>
</tr>
</tbody>
</table>

The output_type attribute is not mandatory. If not specified, then CLOB is the default.

**rowid_parameter**
When you specify TRUE, the rowid of the document to be filtered is passed as the first parameter, before the input and output parameters.

For example, with INPUT_TYPE BLOB, OUTPUT_TYPE CLOB, and ROWID_PARAMETER TRUE, the filter procedure must have the signature as follows:
procedure(in rowid, in blob, in out nocopy clob)

This attribute is useful for when your procedure requires data from other columns or tables. This attribute is not mandatory. The default is FALSE.

**format_parameter**
When you specify TRUE, the value of the format column of the document being filtered is passed to the filter procedure before input and output parameters, but after the rowid parameter, if enabled.

Specify the name of the format column at index time in the parameters string, using the keyword `format column <columnname>` . The parameter type must be IN VARCHAR2.

The format column value can be read by means of the rowid parameter, but this attribute enables a single filter to work on multiple table structures, because the format attribute is abstracted and does not require the knowledge of the name of the table or format column.

FORMAT_PARAMETER is not mandatory. The default is FALSE.

**charset_parameter**
When you specify TRUE, the value of the charset column of the document being filtered is passed to the filter procedure before input and output parameters, but after the rowid and format parameter, if enabled.

Specify the name of the charset column at index time in the parameters string, using the keyword `charset column <columnname>` . The parameter type must be IN VARCHAR2.

CHARSET_PARAMETER attribute is not mandatory. The default is FALSE.

**Parameter Order**

ROWID_PARAMETER, FORMAT_PARAMETER, and CHARSET_PARAMETER are all independent. The order is rowid, the format, then charset. However, the filter procedure is passed only the minimum parameters required.

For example, assume that INPUT_TYPE is BLOB and OUTPUT_TYPE is CLOB. If your filter procedure requires all parameters, then the procedure signature must be:

(id IN ROWID, format IN VARCHAR2, charset IN VARCHAR2, input IN BLOB, output IN OUT NOCOPY CLOB)

If your procedure requires only the ROWID, then the procedure signature must be:

(id IN ROWID, input IN BLOB, output IN OUT NOCOPY CLOB)

**Procedure Filter Execute Requirements**

To create an index using a PROCEDURE_FILTER preference, the index owner must have execute permission on the procedure.

**Error Handling**

The filter procedure can raise any errors needed through the normal PL/SQL raise_application_error facility. These errors are propagated to the CTX_USER_INDEX_ERRORS view or reported to the user, depending on how the filter is invoked.
Procedure Filter Preference Example
Consider a filter procedure `CTXSYS.NORMALIZE` that you define with the following signature:

```sql
PROCEDURE NORMALIZE(id IN ROWID, charset IN VARCHAR2, input IN CLOB,
output IN OUT NOCOPY VARCHAR2);
```

To use this procedure as your filter, set up your filter preference as follows:

```sql
begin
ctx_ddl.create_preference('myfilt', 'procedure_filter');
ctx_ddl.set_attribute('myfilt', 'procedure', 'normalize');
ctx_ddl.set_attribute('myfilt', 'input_type', 'clob');
ctx_ddl.set_attribute('myfilt', 'output_type', 'varchar2');
ctx_ddl.set_attribute('myfilt', 'rowid_parameter', 'TRUE');
ctx_ddl.set_attribute('myfilt', 'charset_parameter', 'TRUE');
end;
```

Lexer Types
Use the lexer preference to specify the language of the text to be indexed. To create a lexer preference, you must use one of the lexer types described in Table 2-15.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTO_LEXER</td>
<td>Lexer for indexing columns that contain documents of different languages.</td>
</tr>
<tr>
<td>BASIC_LEXER</td>
<td>Lexer for extracting tokens from text in languages, such as English and most western European languages that use white space delimited words.</td>
</tr>
<tr>
<td>MULTI_LEXER</td>
<td>Lexer for indexing tables containing documents of different languages such as English, German, and Japanese.</td>
</tr>
<tr>
<td>CHINESE_VGRAM_LEXER</td>
<td>Lexer for extracting tokens from Chinese text.</td>
</tr>
<tr>
<td>CHINESE_LEXER</td>
<td>Lexer for extracting tokens from Chinese text. This lexer offers benefits over the CHINESE_VGRAM lexer:</td>
</tr>
<tr>
<td></td>
<td>■ Generates a smaller index</td>
</tr>
<tr>
<td></td>
<td>■ Better query response time</td>
</tr>
<tr>
<td></td>
<td>■ Generates real world tokens resulting in better query precision</td>
</tr>
<tr>
<td></td>
<td>■ Supports stop words</td>
</tr>
<tr>
<td>JAPANESE_VGRAM_LEXER</td>
<td>Lexer for extracting tokens from Japanese text.</td>
</tr>
<tr>
<td>JAPANESE_LEXER</td>
<td>Lexer for extracting tokens from Japanese text. This lexer offers the following advantages over the JAPANESE_VGRAM lexer:</td>
</tr>
<tr>
<td></td>
<td>■ Generates smaller index</td>
</tr>
<tr>
<td></td>
<td>■ Better query response time</td>
</tr>
<tr>
<td></td>
<td>■ Generates real world tokens resulting in better precision</td>
</tr>
<tr>
<td>KOREAN_MORPH_LEXER</td>
<td>Lexer for extracting tokens from Korean text.</td>
</tr>
<tr>
<td>USER_LEXER</td>
<td>Lexer you create to index a particular language.</td>
</tr>
<tr>
<td>WORLD_LEXER</td>
<td>Lexer for indexing tables containing documents of different languages; autodetects languages in a document.</td>
</tr>
</tbody>
</table>
AUTO_LEXER

Use the AUTO_LEXER type to index columns that contain documents of different languages. It performs language identification, word segmentation, document analysis, part-of-speech tagging, and stemming. The AUTO_LEXER also enables customization of these components. Although parts-of-speech information that is generated by the AUTO_LEXER is not exposed for your use, AUTO_LEXER uses it for context-sensitive or tagged stemming.

At index time, AUTO_LEXER automatically detects the language of the document, and tokenizes and stems the document appropriately. At query time, the language of the query is inherited from the query template. If the query template is not used, or if no language is specified in the query template, then the language of the query is inherited from the session language. Table 2–16 lists the supported languages.

See Also:  Appendix I, “AUTO_LEXER Parts-of-Speech Tagging”

Table 2–16  Languages Supported for AUTO_LEXER

<table>
<thead>
<tr>
<th>Language</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARABIC</td>
<td>JAPANESE</td>
</tr>
<tr>
<td>CATALAN</td>
<td>KOREAN</td>
</tr>
<tr>
<td>SIMPLIFIED CHINESE</td>
<td>TRADITIONAL CHINESE</td>
</tr>
<tr>
<td>CROATION</td>
<td>POLISH</td>
</tr>
<tr>
<td>CZECH</td>
<td>PORTUGUESE</td>
</tr>
<tr>
<td>DANISH</td>
<td>ROMANIAN</td>
</tr>
<tr>
<td>DUTCH</td>
<td>RUSSIAN</td>
</tr>
<tr>
<td>ENGLISH</td>
<td>SERBIAN</td>
</tr>
<tr>
<td>FINNISH</td>
<td>SLOVAK</td>
</tr>
<tr>
<td>FRENCH</td>
<td>SLOVENIAN</td>
</tr>
<tr>
<td>GERMAN</td>
<td>SPANISH</td>
</tr>
<tr>
<td>GREEK</td>
<td>SWEDISH</td>
</tr>
<tr>
<td>HEBREW</td>
<td>THAI</td>
</tr>
<tr>
<td>HUNGARIAN</td>
<td>TURKISH</td>
</tr>
<tr>
<td>ITALIAN</td>
<td>NORWEGIAN: NYNORSK</td>
</tr>
<tr>
<td>NORWEGIAN: BOKMAL</td>
<td>PERSIAN</td>
</tr>
</tbody>
</table>

AUTO_LEXER Attributes Inherited from BASIC_LEXER

The following attributes are used in the same way and have the same effect on the AUTO_LEXER as their corresponding attributes in BASIC_LEXER:

- base_letter
- base_letter_type
- override_base_letter
- mixed_case
- alternate_spelling

See Also:  "BASIC_LEXER" on page 2-39
### Table 2–17  AUTO_LEXER Language-Independent Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>language</td>
<td>&lt;characters&gt;</td>
<td>Specifies the possible languages of the input documents. If no language is specified, then AUTO_LEXER performs auto detection.</td>
</tr>
<tr>
<td></td>
<td>(space-delimited string)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>If one language is specified, then the language is set manually and AUTO_LEXER does not perform auto detection.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>If more than one language is specified, then AUTO_LEXER performs auto detection but limits the detected language to be among the language set.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note:</strong> The automatic detection of language is statistically based and, thus, inherently imperfect.</td>
</tr>
<tr>
<td>deriv_stems</td>
<td>NO (disabled)</td>
<td>Specifies whether the derivational stemming should be used or not. Currently, derivational stemming is only available for English. Hence, the DERIV_STEMS has no effect in other languages.</td>
</tr>
<tr>
<td></td>
<td>YES (default)</td>
<td>Also, when derivational stemming is performed, tagging and tag stemming is not used. As a result, the tagging and tagged stemming client dictionary has no effect on the stemming result.</td>
</tr>
<tr>
<td>german_decompound</td>
<td>NO (disabled)</td>
<td>Specifies whether German de-compounding should be performed in the stemmer or not.</td>
</tr>
<tr>
<td></td>
<td>YES (default, enabled for German only)</td>
<td></td>
</tr>
<tr>
<td>sentence_token_limit</td>
<td>n (default 100)</td>
<td>Specifies the maximum number of tokens allowed in a single sentence. This parameter can affect memory usage for AUTO_LEXER. If set too high, then it might cause memory overload. If set too low, then the sentence might be truncated incorrectly.</td>
</tr>
<tr>
<td>index_stems</td>
<td>NO (disabled)</td>
<td>Specifies whether an index stemmer should be used.</td>
</tr>
<tr>
<td></td>
<td>YES (default)</td>
<td>If specified as YES, then the stemmer that corresponds to the document language will be used and the stemmer will always be configured to maximize document recall.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>If specified as NO, then queries with stem operators will use the word list stemming to try to stem the tokens. If word list stemming is not available, then the stem operator will be ignored.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>For documents in Finnish, Swedish, Dutch, if the index_stems is set to YES, then compound word stemming will automatically be performed, and compounds are always separated into their component stems.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note:</strong> If INDEX_STEMS attribute is set to YES, then the STEMMER attribute of BASIC_WORDLIST will be ignored, and the stemmer used by the auto_lexer will be used during query to determine the stem of the given query term.</td>
</tr>
<tr>
<td></td>
<td>YES (default)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2–18 AUTO_LEXER Language-Dependent Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>base_letter</td>
<td>NO (disabled)</td>
<td>Specify whether characters that have diacritical marks (umlauts, cedillas, acute accents, and so on) are converted to their base form before being stored in the Text index.</td>
</tr>
<tr>
<td></td>
<td>YES (enabled)</td>
<td></td>
</tr>
<tr>
<td>base_letter_type</td>
<td>GENERIC (default)</td>
<td>The GENERIC value is the default and means that base letter transformation uses one transformation table that applies to all languages.</td>
</tr>
<tr>
<td></td>
<td>SPECIFIC</td>
<td></td>
</tr>
<tr>
<td>override_base_letter</td>
<td>TRUE</td>
<td>When base_letter is enabled at the same time as alternate_spelling, it is sometimes necessary to override base_letter to prevent unexpected results from serial transformations.</td>
</tr>
<tr>
<td></td>
<td>FALSE (default)</td>
<td></td>
</tr>
<tr>
<td>mixed_case</td>
<td>NO (disabled)</td>
<td>Specify whether the lexer leaves the tokens exactly as they appear in the text or converts the tokens to all uppercase. The default is NO (tokens are converted to all uppercase).</td>
</tr>
<tr>
<td></td>
<td>YES (enabled)</td>
<td></td>
</tr>
<tr>
<td>alternate_spelling</td>
<td>GERMAN (German alternate spelling)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DANISH (Danish alternate spelling)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SWEDISH (Swedish alternate spelling)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NONE (No alternate spelling, default)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2–17 (Cont.) AUTO_LEXER Language-Dependent Attributes

Table 2–18 AUTO_LEXER Language-Dependent Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;language&gt;_prefix_morphemes</td>
<td>characters (space-delimited string)</td>
<td>Specifies the list of inflectional prefixes that, when enclosed by parentheses, are kept together with the base word. For example, (re) analyze.</td>
</tr>
<tr>
<td>&lt;language&gt;_suffix_morphemes</td>
<td>characters (space-delimited string)</td>
<td>Specifies the list of inflectional suffixes that, when enclosed by parentheses are kept together with the base word. For example, file(s).</td>
</tr>
<tr>
<td>&lt;language&gt;_punctuations</td>
<td>characters (space-delimited string)</td>
<td>Specifies punctuation that breaks sentences.</td>
</tr>
<tr>
<td>&lt;language&gt;_sentence_starts</td>
<td>characters (space-delimited string)</td>
<td>Specifies text that starts a sentence.</td>
</tr>
<tr>
<td>&lt;language&gt;_non_sentence_ending_abrev</td>
<td>characters (space-delimited string)</td>
<td>Specifies abbreviations that do not end sentences.</td>
</tr>
</tbody>
</table>
### Table 2–19  Default Values for AUTO_LEXER Language-Dependent Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Language</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;language&gt;_prefix_morphemes</code></td>
<td>All languages</td>
<td>None</td>
</tr>
<tr>
<td><code>&lt;language&gt;_suffix_morphemes</code></td>
<td>English</td>
<td>s es er</td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
<td>ba n s es</td>
</tr>
<tr>
<td></td>
<td>Portuguese</td>
<td>s es</td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>in innen</td>
</tr>
<tr>
<td></td>
<td>French</td>
<td>ne e</td>
</tr>
<tr>
<td></td>
<td>All other languages</td>
<td>None</td>
</tr>
<tr>
<td><code>&lt;language&gt;_punctuations</code></td>
<td>English</td>
<td>. ? !</td>
</tr>
<tr>
<td></td>
<td>Arabic, Catalan, Croatian,</td>
<td>. ? ! --</td>
</tr>
<tr>
<td></td>
<td>Czech, English, Greek,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hebrew, Hungarian,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Polish, Romanian,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Russian, Serbian, Slovak,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slovenian, Turkish</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bokmal, Danish, Finnish,</td>
<td>, ? !</td>
</tr>
<tr>
<td></td>
<td>French, German, Italian,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Korean, Nynorsk,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Portuguese, Spanish,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Swedish</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Japanese</td>
<td>. ? ! . ? ! :</td>
</tr>
<tr>
<td></td>
<td>Persian</td>
<td>. ! !</td>
</tr>
<tr>
<td><code>&lt;language&gt;_sentence_starts</code></td>
<td>Arabic, catalan, Croatian,</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>Czech, English, Finish,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>French, Greek, Hebrew,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hungarian, Italian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Japanese, Korean, Persian,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Polish, Portuguese,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Romanian, Russian,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Serbian, Simplified Chinese,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slovak, Slovenian, Thai,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Traditional Chinese,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Turkish</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bokmal, Danish, Nynorsk,</td>
<td>&quot;-</td>
</tr>
<tr>
<td></td>
<td>Swedish</td>
<td></td>
</tr>
<tr>
<td></td>
<td>German</td>
<td>&quot; &quot; &quot;</td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
<td>&quot; &quot; i</td>
</tr>
<tr>
<td><code>&lt;language&gt;_non_sentence_ending_abrev</code></td>
<td>Arabic, Polish, Romanian, e.g. i.e. viz. a.k.a.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Russian, Serbian, Slovak,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slovenian, Turkish</td>
<td></td>
</tr>
<tr>
<td>Attribute</td>
<td>Language</td>
<td>Default Value</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>sekr. førstekons. førstesekr. førsteaman.</td>
<td>fullm. frk. d.e. d.y. dr. dir. aman. adm.dir. adm. dir.</td>
</tr>
<tr>
<td>Croatian, Czech,</td>
<td>e.g. i.e. viz. a.k.a.</td>
<td></td>
</tr>
<tr>
<td>Greek, Hebrew,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>eq. d.e d.ae. d.y. dr. dir.</td>
<td>dept.chef civiling, bibl. ass. admin. adj. Skt. H.K.H.</td>
</tr>
<tr>
<td>Finish</td>
<td>inkl. dipl. prof. hrr. hr. Hrr. Dr. dr.</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>c.-à-d. cf. e.g. ex. i.e Pr. Prof. M. Mr.</td>
<td>Mrs. Mme Mmes Mile Milles Mgr. MM. Lieut. Gén. Dr. Col.</td>
</tr>
<tr>
<td>German</td>
<td>ca. bzw. e.g. i.e. inkl. Fr. Frl. Mme. Mile</td>
<td>Mag. Stud. Tel. Hr. Hrn. ap.Prof. Prof.</td>
</tr>
<tr>
<td>Italian</td>
<td>e.g. i.e. pag. pagg. tel. T.V. N.H. N.D.</td>
<td>comm. col. cav. cap. geom. gen. ing. jr. mr. mons. mar. magg. prof. prof.</td>
</tr>
<tr>
<td></td>
<td>vesc. S.S. S.E. avv. app. amm. arch. on.</td>
<td>dir. dott. dott.ssa dr. rag.</td>
</tr>
<tr>
<td>Korean</td>
<td>e.g. i.e. a.k.a. Dr. Mr. Mrs. Ms. Prof.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gen.sekr. førstekons. førstesekr. førsteaman.</td>
<td>fullm. frk. d.e. d.y. dr. dir. aman. adm.dir. adm. dir. fyrstesekr.</td>
</tr>
<tr>
<td></td>
<td>fyrstekons. fyrsteaman. hr</td>
<td></td>
</tr>
<tr>
<td>Persian</td>
<td>Dr. Mr. Mrs. Ms. Prof. e.g. i.e. viz. a.k.a.</td>
<td></td>
</tr>
<tr>
<td>Swedish</td>
<td>inkl. prof. hrr. hr. Hrr. Dr. dr. Dr.</td>
<td></td>
</tr>
</tbody>
</table>
Examples for AUTO_LEXER Language-Dependent Attributes

Example 2–1  <language>_prefix_morphemes
ctxt_ddl.set_attribute(
    'a_lex', 'english_prefix_morphemes', 're'
);

Example 2–2  <language>_suffix_morphemes
ctxt_ddl.set_attribute(
    'a_lex', 'english_suffix_morphemes', 's es'
);

Example 2–3  <language>_punctuations
ctxt_ddl.set_attribute(
    'a_lex', 'english_punctuations', '. ? !'
);

Example 2–4  <language>_sentence_starts
ctxt_ddl.set_attribute(
    'a_lex', 'english_sentence_starts', '" '
);

Example 2–5  <language>_non_sentence_ending_abbrev
ctxt_ddl.set_attribute(
    'a_lex', 'english_non_sentence_ending_abbrev', 'e.g. a.k.a. Dr.'
);

AUTO_LEXER User-Defined Dictionary Attributes

The attributes in this section are language-specific and are used to set the name of user-supplied dictionary files. The attributes share the following behavior:

- The value of the attribute specifies only the filename (excluding the file path) of the dictionary. The file should be placed at the following location: $ORACLE_HOME/ctx/data/user.
- The set_attribute method does not load the file; it only records the filename. Therefore, the file must be at the specified location when the dictionary is needed. Otherwise, an error will be raised.
- The client dictionaries specify the character encoding of the file in the <?xml ...?> markup (for example, <?xml encoding="cp-1252" ?>). If no <?xml ...?> specification exists and no special identification is given (for example, as for UCS-4), then the system assumes that the encoding is cp-1252, instead of UTF-8 as specified in the XML standard.

The user-supplied dictionary attributes are listed in Table 2–20, and examples are given for each attribute under “Examples for AUTO_LEXER User-Supplied Dictionary Attributes” on page 2-37. The values inside the brackets in Table 2–20 refer to parts-of-speech tags.

See Also: Appendix I, "AUTO_LEXER Parts-of-Speech Tagging" for the detailed list of parts-of-speech tags for each supported language.
Table 2–20  AUTO_LEXER User-Supplied Dictionary Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>language_abbrev_dictionary</td>
<td>Any valid filename.</td>
<td>Specifies the file name of the user-supplied abbreviation dictionary. The abbreviation dictionary lets you add custom abbreviations that should be processed as such. The abbreviation dictionary is used in word segmentation and document analysis to help resolve ambiguity and, therefore, improve sentence/paragraph searches.</td>
</tr>
<tr>
<td>language_tagging_dictionary</td>
<td>Any valid filename.</td>
<td>Specifies the file name of the user-supplied tagging dictionary. The tagging dictionary lets you add the appropriate part-of-speech tag for words that may not occur in the prepackaged tagging lexicons. The objective of setting this attribute is to improve internal parts-of-speech tagging and, therefore, improve the quality of context-sensitive stemming. This is accomplished by providing higher quality context using the POS information of the surrounding words.</td>
</tr>
<tr>
<td>language_tagged_stemming_dictionary</td>
<td>Any valid filename.</td>
<td>Specifies the file name of the user-supplied tagged stemming dictionary. The tagged stemming dictionary lets you add the appropriate stem for words of a particular part of speech that may not occur in the prepackaged tagged stemming lexicons.</td>
</tr>
<tr>
<td>language_stemming_dictionary</td>
<td>Any valid filename.</td>
<td>Specifies the file name of the user-supplied stemming dictionary. The stemming dictionary lets you add the appropriate stem for words that may not occur in the prepackaged tagged stemming lexicons.</td>
</tr>
<tr>
<td>language_CCJT_dictionary</td>
<td>Any valid filename.</td>
<td>Specifies the file name of the user-supplied CCJT dictionary. The CCJT dictionaries can be used for each of the languages in a limited way to influence all the following processes: segmentation, tagging, stemming and tagged stemming. (As Thai supports only segmentation and stemming, the client dictionary cannot be used for tagging.) The processing algorithm for the CCJT languages uses pre-calculated probabilities of tag sequences to segment, stem and tag a sentence. For this reason, it cannot be guaranteed that a user-defined entry will always be returned for all parts of speech. However, client dictionary entries tagged Nn or Nn-Prop will be given precedence and will always be returned in the final results.</td>
</tr>
</tbody>
</table>

Table 2–21  Languages and Default Values for AUTO_LEXER User-Supplied Dictionary Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Language</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>language_ abbrev_dictionary</td>
<td>All</td>
<td>Null</td>
</tr>
<tr>
<td>language_tagging_dictionary</td>
<td>All</td>
<td>Null</td>
</tr>
<tr>
<td>language_tagged_stemming_dictionary</td>
<td>All</td>
<td>Null</td>
</tr>
<tr>
<td>language_stemming_dictionary</td>
<td>All</td>
<td>Null</td>
</tr>
<tr>
<td>language_CCJT_dictionary</td>
<td>All</td>
<td>Null</td>
</tr>
</tbody>
</table>

Examples for AUTO_LEXER User-Supplied Dictionary Attributes

**Example 2–6  <language>_abbrev_dictionary**

```bash
eexec Ctx_ddl.set_attribute {
    'a_lex',
    'english abbrev dictionary',
    'english abbrev dictionary.xml'
};
```
Dictionary example:

```xml
<explicit-pair-list>
  <item key="inc." analysis="[Abbrev]"/>
  <item key="Inc." analysis="[Abbrev]"/>
</explicit-pair-list>
```

**Example 2–7** `<language>_tagging_dictionary`

```sql
exec Ctx_ddl.set_attribute
     ('a_lex', 'english_tagging_dictionary', 'english_tagging_dictionary.xml');
```

Dictionary example:

```xml
<explicit-pair-list>
  <item key="Inxight" analysis = "[Prop]"> </item>
  <item key="furby" analysis = "[Nn-Indef-Sg]"> </item>
  <item key="furbys" analysis = "[Nn-Indef-Sg-Gen]"> </item>
  <item key="furbyer" analysis = "[Nn-Indef-Pl]"> </item>
  <item key="furbyen" analysis = "[Nn-Def-Sg]"> </item>
  <item key="furbyene" analysis = "[Nn-Def-Pl]"> </item>
</explicit-pair-list>
```

The values inside the brackets are part-of-speech tags.

**Example 2–8** `<language>_tagged_stemming_dictionary`

```sql
exec Ctx_ddl.set_attribute {
  'a_lex',
  'english_tagged_stemming_dictionary',
  'english_tagged_stem_dictionary.xml'
};
```

Dictionary example:

```xml
<tag-stem-list>
  <item key = "running[V-PrPart]" stem = "run" > </item>
</tag-stem-list>
```

The values inside the brackets are part-of-speech tags.

**Example 2–9** `<language>_stemming_dictionary`

```sql
exec Ctx_ddl.set_attribute {
  'a_lex',
  'english_stemming_dictionary',
  'english_stemming_dictionary.xml'
};
```

Dictionary example:

```xml
<explicit-pair-list>
  <item key="running" stem="run"></item>
  <item key="flying" stem="fly"></item>
</explicit-pair-list>
```

**Example 2–10** `<language>_CCJT_dictionary`

```sql
exec Ctx_ddl.set_attribute {
```
BASIC_LEXER

Use the BASIC_LEXER type to identify tokens for creating Text indexes for English and all other supported whitespace-delimited languages.

The BASIC_LEXER also enables base-letter conversion, composite word indexing, case-sensitive indexing and alternate spelling for whitespace-delimited languages that have extended character sets.

In English and French, you can use the BASIC_LEXER to enable theme indexing.

---

**Note:** Any processing that the lexer does to tokens before indexing (for example, removal of characters, and base-letter conversion) are also performed on query terms at query time. This ensures that the query terms match the form of the tokens in the Text index.

---

BASIC_LEXER supports any database character set.

BASIC_LEXER has the attributes shown in Table 2–22.

### Table 2–22  BASIC_LEXER Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>continuation</td>
<td>characters</td>
</tr>
<tr>
<td>numgroup</td>
<td>characters</td>
</tr>
<tr>
<td>numjoin</td>
<td>characters</td>
</tr>
<tr>
<td>printjoins</td>
<td>characters</td>
</tr>
<tr>
<td>punctuations</td>
<td>characters</td>
</tr>
<tr>
<td>skipjoins</td>
<td>characters</td>
</tr>
<tr>
<td>startjoins</td>
<td>non alphanumeric characters that occur at the beginning of a token (string)</td>
</tr>
<tr>
<td>endjoins</td>
<td>non alphanumeric characters that occur at the end of a token (string)</td>
</tr>
<tr>
<td>whitespace</td>
<td>characters (string)</td>
</tr>
<tr>
<td>Attribute</td>
<td>Attribute Value</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------------------------------------</td>
</tr>
<tr>
<td>newline</td>
<td>NEWLINE (\n)</td>
</tr>
<tr>
<td></td>
<td>CARRIAGE_RETURN (\r)</td>
</tr>
<tr>
<td>base_letter</td>
<td>NO (disabled)</td>
</tr>
<tr>
<td></td>
<td>YES (enabled)</td>
</tr>
<tr>
<td>base_letter_type</td>
<td>GENERIC (default)</td>
</tr>
<tr>
<td></td>
<td>SPECIFIC</td>
</tr>
<tr>
<td>override_base_letter</td>
<td>TRUE</td>
</tr>
<tr>
<td></td>
<td>FALSE (default)</td>
</tr>
<tr>
<td>mixed_case</td>
<td>NO (disabled)</td>
</tr>
<tr>
<td></td>
<td>YES (enabled)</td>
</tr>
<tr>
<td>composite</td>
<td>DEFAULT (no composite word indexing, default)</td>
</tr>
<tr>
<td></td>
<td>GERMAN (German composite word indexing)</td>
</tr>
<tr>
<td></td>
<td>DUTCH (Dutch composite word indexing)</td>
</tr>
</tbody>
</table>
Table 2–22  (Cont.) BASIC_LEXER Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>index_stems</td>
<td>0 NONE</td>
</tr>
<tr>
<td></td>
<td>1 ENGLISH</td>
</tr>
<tr>
<td></td>
<td>2 DERIVATIONAL</td>
</tr>
<tr>
<td></td>
<td>3 DUTCH</td>
</tr>
<tr>
<td></td>
<td>4 FRENCH</td>
</tr>
<tr>
<td></td>
<td>5 GERMAN</td>
</tr>
<tr>
<td></td>
<td>6 ITALIAN</td>
</tr>
<tr>
<td></td>
<td>7 SPANISH</td>
</tr>
<tr>
<td></td>
<td>8 ARABIC</td>
</tr>
<tr>
<td></td>
<td>9 BOKMAL</td>
</tr>
<tr>
<td></td>
<td>10 CATALAN</td>
</tr>
<tr>
<td></td>
<td>11 CROATIAN</td>
</tr>
<tr>
<td></td>
<td>12 CZECH</td>
</tr>
<tr>
<td></td>
<td>13 DANISH</td>
</tr>
<tr>
<td></td>
<td>28 DERIVATIONAL_NEW (see Note)</td>
</tr>
<tr>
<td></td>
<td>29 DUTCH_NEW (see Note)</td>
</tr>
<tr>
<td></td>
<td>27 ENGLISH_NEW (see Note)</td>
</tr>
<tr>
<td></td>
<td>14 FINNISH (see Note)</td>
</tr>
<tr>
<td></td>
<td>30 FRENCH_NEW (see Note)</td>
</tr>
<tr>
<td></td>
<td>31 GERMAN_NEW (see Note)</td>
</tr>
<tr>
<td></td>
<td>15 GREEK</td>
</tr>
<tr>
<td></td>
<td>16 HEBREW</td>
</tr>
<tr>
<td></td>
<td>17 HUNGARIAN</td>
</tr>
<tr>
<td></td>
<td>32 ITALIAN_NEW (see Note)</td>
</tr>
<tr>
<td></td>
<td>18 NYNORSK</td>
</tr>
<tr>
<td></td>
<td>19 POLISH</td>
</tr>
<tr>
<td></td>
<td>20 PORTUGUESE</td>
</tr>
<tr>
<td></td>
<td>21 ROMANIAN</td>
</tr>
<tr>
<td></td>
<td>22 RUSSIAN</td>
</tr>
<tr>
<td></td>
<td>23 SERBIAN</td>
</tr>
<tr>
<td></td>
<td>24 SLOVAK</td>
</tr>
<tr>
<td></td>
<td>25 SLOVENIAN</td>
</tr>
<tr>
<td></td>
<td>33 SPANISH_NEW (see Note)</td>
</tr>
<tr>
<td></td>
<td>26 SWEDISH (see Note)</td>
</tr>
</tbody>
</table>

**Note:** De-compounding word stemming is automatically performed when `index_stems` is set to SWEDISH, FINNISH, or DUTCH_NEW values.

**Note:** Seven of the `index_stem` attributes that are new for this release have a "_NEW" suffix to enable you to utilize the new stemmer attributes while maintaining backward compatibility with previous releases of Oracle Text.

<table>
<thead>
<tr>
<th>indexThemes</th>
<th>YES (enabled)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO (disabled, default)</td>
</tr>
</tbody>
</table>
**Table 2–22 (Cont.) BASIC_LEXER Attributes**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>index_text</td>
<td>NO (disabled, default)</td>
</tr>
<tr>
<td></td>
<td>YES (enabled, default)</td>
</tr>
<tr>
<td>prove_themes</td>
<td>NO (disabled)</td>
</tr>
<tr>
<td></td>
<td>YES (enabled, default)</td>
</tr>
<tr>
<td>theme_language</td>
<td>AUTO (default)</td>
</tr>
<tr>
<td></td>
<td>(any Globalization Support language)</td>
</tr>
<tr>
<td>alternate_spelling</td>
<td>GERMAN (German alternate spelling)</td>
</tr>
<tr>
<td></td>
<td>DANISH (Danish alternate spelling)</td>
</tr>
<tr>
<td></td>
<td>SWEDISH (Swedish alternate spelling)</td>
</tr>
<tr>
<td></td>
<td>NONE (No alternate spelling, default)</td>
</tr>
<tr>
<td>new_german_spelling</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>NO (default)</td>
</tr>
</tbody>
</table>

**continuation**
Specify the characters that indicate a word continues on the next line and should be indexed as a single token. The most common continuation characters are hyphen ‘-’ and backslash ‘\’.

**numgroup**
Specify a single character that, when it appears in a string of digits, indicates that the digits are groupings within a larger single unit.

For example, comma ‘,’ might be defined as a numgroup character because it often indicates a grouping of thousands when it appears in a string of digits.

**numjoin**
Specify the characters that, when they appear in a string of digits, cause Oracle Text to index the string of digits as a single unit or word.

For example, period ‘.’ can be defined as numjoin characters because it often serves as decimal points when it appears in a string of digits.

---

**Note:** The default values for numjoin and numgroup are determined by the globalization support initialization parameters that are specified for the database.

In general, a value need not be specified for either numjoin or numgroup when creating a lexer preference for BASIC_LEXER.

---

**printjoins**
Specify the non alphanumeric characters that, when they appear anywhere in a word (beginning, middle, or end), are processed as alphanumeric and included with the token in the Text index. This includes printjoins that occur consecutively.

For example, if the hyphen ‘-’ and underscore ‘_’ characters are defined as printjoins, terms such as *pseudo-intellectual* and *_file_* are stored in the Text index as *pseudo-intellectual* and *_file_*.
punctuations
Specify a list of non-alphanumeric characters that, when they appear at the end of a
word, indicate the end of a sentence. The defaults are period '.', question mark '?', and
exclamation point '!'.

Characters that are defined as punctuations are removed from a token before text
indexing. However, if a punctuations character is also defined as a printjoins
classifier, then the character is removed only when it is the last character in the token.

For example, if the period (.) is defined as both a printjoins and a punctuations
classifier, then the following transformations take place during indexing and querying
as well:

<table>
<thead>
<tr>
<th>Token</th>
<th>Indexed Token</th>
</tr>
</thead>
<tbody>
<tr>
<td>.doc</td>
<td>.doc</td>
</tr>
<tr>
<td>dog.doc</td>
<td>dog.doc</td>
</tr>
<tr>
<td>dog..doc</td>
<td>dog..doc</td>
</tr>
<tr>
<td>dog.</td>
<td>dog</td>
</tr>
<tr>
<td>dog...</td>
<td>dog..</td>
</tr>
</tbody>
</table>

In addition, BASIC_LEXER use punctuations characters in conjunction with
newline and whitespace characters to determine sentence and paragraph
delimiters for sentence/paragraph searching.

skipjoins
Specify the non-alphanumeric characters that, when they appear within a word,
identify the word as a single token; however, the characters are not stored with the
token in the Text index.

For example, if the hyphen character '-' is defined as a skipjoins, then the word
pseudo-intellectual is stored in the Text index as pseudointellectual.

startjoins/endjoins
For startjoins, specify the characters that when encountered as the first character
in a token explicitly identify the start of the token. The character, as well as any other
startjoins characters that immediately follow it, is included in the Text index entry
for the token. In addition, the first startjoins character in a string of startjoins
characters implicitly ends the previous token.

For endjoins, specify the characters that when encountered as the last character in a
token explicitly identify the end of the token. The character, as well as any other
startjoins characters that immediately follow it, is included in the Text index entry
for the token.

Note: Printjoins and skipjoins are mutually exclusive. The
same characters cannot be specified for both attributes.
The following rules apply to both startjoins and endjoins:

- The characters specified for startjoins/endjoins cannot occur in any of the other attributes for BASIC_LEXER.
- startjoins/endjoins characters can occur only at the beginning or end of tokens

Printjoins differ from endjoins and startjoins in that position does not matter. For example, $35 will be indexed as one token if $ is a startjoin or a printjoin, but as two tokens if it is defined as an endjoin.

**whitespace**
Specify the characters that are treated as blank spaces between tokens. BASIC_LEXER uses whitespace characters in conjunction with punctuations and newline characters to identify character strings that serve as sentence delimiters for sentence and paragraph searching.

The predefined default values for whitespace are space and tab. These values cannot be changed. Specifying characters as whitespace characters adds to these defaults.

**newline**
Specify the characters that indicate the end of a line of text. BASIC_LEXER uses newline characters in conjunction with punctuations and whitespace characters to identify character strings that serve as paragraph delimiters for sentence and paragraph searching.

The only valid values for newline are NEWLINE and CARRIAGE_RETURN (for carriage returns). The default is NEWLINE.

**base_letter**
Specify whether characters that have diacritical marks (umlauts, cedillas, acute accents, and so on) are converted to their base form before being stored in the Text index. The default is NO (base-letter conversion disabled). For more information on base-letter conversions and base_letter_type, see Base-Letter Conversion on page 16-2.

**base_letter_type**
Specify GENERIC or SPECIFIC.

The GENERIC value is the default and means that base letter transformation uses one transformation table that applies to all languages. For more information on base-letter conversions and base_letter_type, see "Base-Letter Conversion" on page 16-2.

**override_base_letter**
When base_letter is enabled at the same time as alternate_spelling, it is sometimes necessary to override base_letter to prevent unexpected results from serial transformations. See "Overriding Base-Letter Transformations with Alternate Spelling" on page 16-3. Default is FALSE.

**mixed_case**
Specify whether the lexer leaves the tokens exactly as they appear in the text or converts the tokens to all uppercase. The default is NO (tokens are converted to all uppercase).
composite
Specify whether composite word indexing is disabled or enabled for either GERMAN or DUTCH text. The default is DEFAULT (composite word indexing disabled).

Words that are usually one entry in a German dictionary are not split into composite stems, while words that aren't dictionary entries are split into composite stems.

To retrieve the indexed composite stems, you must enter a stem query, such as $bahnhof. The language of the wordlist stemmer must match the language of the composite stems.

Stemming User-Dictionaries
You can create a user-dictionary for your own language to customize how words are decomposed. These dictionaries are shown in Table 2–23.

Table 2–23  Stemming User-Dictionaries

<table>
<thead>
<tr>
<th>Dictionary</th>
<th>Stemmer</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ORACLE_HOME/ctx/data/frlx/drfr.dct</td>
<td>French</td>
</tr>
<tr>
<td>$ORACLE_HOME/ctx/data/delx/drde.dct</td>
<td>German</td>
</tr>
<tr>
<td>$ORACLE_HOME/ctx/data/nllx/drnl.dct</td>
<td>Dutch</td>
</tr>
<tr>
<td>$ORACLE_HOME/ctx/data/itlx/drit.dct</td>
<td>Italian</td>
</tr>
<tr>
<td>$ORACLE_HOME/ctx/data/eslx/dres.dct</td>
<td>Spanish</td>
</tr>
<tr>
<td>$ORACLE_HOME/ctx/data/enlx/dren.dct</td>
<td>English and Derivational</td>
</tr>
</tbody>
</table>

Stemming user-dictionaries are not supported for languages other than those listed in Table 2–23.

The format for the user dictionary is as follows:

output term <tab> input term

The individual parts of the decomposed word must be separated by the # character.

The following example entries are for the German word Hauptbahnhof:

Hauptbahnhof<tab>Haupt#Bahnhof
Hauptbahnhof<tab>Haupt#Bahnhof
Hauptbahnhof<tab>Haupt#Bahnhof
Hauptbahnhof<tab>Haupt#Bahnhof
Hauptbahnhof<tab>Haupt#Bahnhof

index_themes
Specify YES to index theme information in English or French. This makes ABOUT queries more precise. The index_themes and index_text attributes cannot both be NO.

If you use the BASIC_LEXER and specify no value for index_themes, this attribute defaults to NO.

You can set this parameter to TRUE for any index type, including CTXCAT. To enter an ABOUT query with CATSEARCH, use the query template with CONTEXT grammar.

Note: Oracle Text ensures that word queries match the case sensitivity of the index being queried. As a result, if you enable case sensitivity for your Text index, queries against the index are always case sensitive.
prove_themes
Specify YES to prove themes. Theme proving attempts to find related themes in a
document. When no related themes are found, parent themes are eliminated from the
document.

While theme proving is acceptable for large documents, short text descriptions with a
few words rarely prove parent themes, resulting in poor recall performance with
ABOUT queries.

Theme proving results in higher precision and less recall (less rows returned) for
ABOUT queries. For higher recall in ABOUT queries and possibly less precision, you can
disable theme proving. Default is YES.

The prove_themes attribute is supported for CONTEXT and CTXRULE indexes.

theme_language
Specify which knowledge base to use for theme generation when index_themes is
set to YES. When index_themes is NO, setting this parameter has no effect on
anything.

Specify any globalization support language or AUTO. You must have a knowledge base
for the language you specify. This release provides a knowledge base in only English
and French. In other languages, you can create your own knowledge base.

See Also: "Adding a Language-Specific Knowledge Base" on
page 15-9 in Chapter 15, "Oracle Text Utilities".

The default is AUTO, which instructs the system to set this parameter according to the
language of the environment.

index_stems
Specify the stemmer to use for stem indexing. Choose one of the following stemmers:

- NONE
- ARABIC
- CATALAN
- CROATIAN
- CZECH
- DANISH
- DERIVATIONAL
- DUTCH
- ENGLISH
- FINNISH
- FRENCH
- GERMAN
- HEBREW
- HUNGARIAN
- ITALIAN
- NORWEGIAN
- POLISH
- PORTUGUESE
- ROMANIAN
- SLOVAK
- SLOVENIAN
- SPANISH
- SWEDISH
Tokens are stemmed to a single base form at index time in addition to the normal forms. Indexing stems enables better query performance for stem ($) queries, such as $computed.

**Note:** If the index_stems attribute is set to one of the languages with ID 8 to 33, which are listed Table 2–22, "BASIC_LEXER Attributes", then the stemmer attribute of BASIC_WORDLIST will be ignored and the stemmer used by the BASIC_LEXER will be used during query to determine the stem of the given query term.

**index_text**
Specify YES to index word information. The index_themes and index_text attributes cannot both be NO.

The default is NO.

**alternate_spelling**
Specify either GERMAN, DANISH, or SWEDISH to enable the alternate spelling in one of these languages. Enabling alternate spelling enables you to query a word in any of its alternate forms.

Alternate spelling is off by default; however, in the language-specific scripts that Oracle provides in admin/defaults (drdefd.sql for German, drdefdk.sql for Danish, and drdefs.sql for Swedish), alternate spelling is turned on. If your installation uses these scripts, then alternate spelling is on. However, you can specify NONE for no alternate spelling. For more information about the alternate spelling conventions Oracle Text uses, see Alternate Spelling on page 16-2.

**new_german_spelling**
Specify whether the queries using the BASIC_LEXER return both traditional and reformed (new) spellings of German words. If new_german_spelling is set to YES, then both traditional and new forms of words are indexed. If it is set to NO, then the word will be indexed only as it as provided in the query. The default is NO.

See Also: "New German Spelling" on page 16-2

**BASIC_LEXER Example**
The following example sets printjoin characters and disables theme indexing with the BASIC_LEXER:

```sql
begin
ctx_ddl.create_preference('mylex', 'BASIC_LEXER');
ctx_ddl.set_attribute('mylex', 'printjoins', '_-');
ctx_ddl.set_attribute ('mylex', 'index_themes', 'NO');
ctx_ddl.set_attribute ( 'mylex', 'index_text', 'YES');
end;
```

To create the index with no theme indexing and with printjoin characters set as described, enter the following statement:

```sql
create index myindex on mytable ( docs )
  indextype is ctxsys.context
  parameters ( 'LEXER mylex' );
```
MULTI_LEXER

Use MULTI_LEXER to index text columns that contain documents of different languages. For example, use this lexer to index a text column that stores English, German, and Japanese documents.

This lexer has no attributes.

You must have a language column in your base table. To index multi-language tables, specify the language column when you create the index.

Create a multi-lexer preference with CTX_DDL.CREATE_PREFERENCE. Add language-specific lexers to the multi-lexer preference with the CTX_DDL.ADD_SUB_LEXER procedure.

During indexing, the MULTI_LEXER examines each row’s language column value and switches in the language-specific lexer to process the document.

The WORLD_LEXER lexer also performs multi-language indexing, but without the need for separate language columns (that is, it has automatic language detection). For more on WORLD_LEXER, see "WORLD_LEXER" on page 2-67.

Multi-language Stoplists

When you use the MULTI_LEXER, you can also use a multi-language stoplist for indexing.

See Also: "Multi-Language Stoplists" on page 2-83.

MULTI_LEXER Example

Create the multi-language table with a primary key, a text column, and a language column as follows:

create table globaldoc {
    doc_id number primary key,
    lang varchar2(3),
    text clob
};

Assume that the table holds mostly English documents, with the occasional German or Japanese document. To handle the three languages, you must create three sub-lexers, one for English, one for German, and one for Japanese:

ctx_ddl.create_preference('english_lexer','basic_lexer');
ctx_ddl.set_attribute('english_lexer','index_themes','yes');
ctx_ddl.set_attribute('english_lexer','theme_language','english');

ctx_ddl.create_preference('german_lexer','basic_lexer');
ctx_ddl.set_attribute('german_lexer','composite','german');
ctx_ddl.set_attribute('german_lexer','mixed_case','yes');
ctx_ddl.set_attribute('german_lexer','alternate_spelling','german');

ctx_ddl.create_preference('japanese_lexer','japanese_vgram_lexer');

Create the multi-lexer preference:

ctx_ddl.create_preference('global_lexer', 'multi_lexer');

Because the stored documents are mostly English, make the English lexer the default using CTX_DDL.ADD_SUB_LEXER:

ctx_ddl.add_sub_lexer('global_lexer','default','english_lexer');
Now add the German and Japanese lexers in their respective languages with CTX_DDL.ADD_SUB_LEXER procedure. Also assume that the language column is expressed in the standard ISO 639-2 language codes, so add those as alternative values.

```
ctx_ddl.add_sub_lexer('global_lexer','german','german_lexer','ger');
ctx_ddl.add_sub_lexer('global_lexer','japanese','japanese_lexer','jpn');
```

Now create the index `globalx`, specifying the multi-lexer preference and the language column in the parameter clause as follows:

```
create index globalx on globaldoc(text) indextype is ctxsys.context
parameters ('lexer global_lexer language column lang');
```

**Querying Multi-Language Tables**

At query time, the multi-lexer examines the language setting and uses the sub-lexer preference for that language to parse the query.

If the language is not set, then the default lexer is used. Otherwise, the query is parsed and run as usual. The index contains tokens from multiple languages, so such a query can return documents in several languages. To limit your query to a given language, use a structured clause on the language column.

If the language column is set to `AUTO`, then the multi-lexer detects the language of the document for the supported languages shown in Table 2–24.

**Table 2–24 Languages Supported for MULTI_LEXER Auto-detection**

<table>
<thead>
<tr>
<th>Language</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARABIC</td>
<td>JAPANESE</td>
</tr>
<tr>
<td>CATALAN</td>
<td>KOREAN</td>
</tr>
<tr>
<td>TRADITIONAL CHINESE</td>
<td>NORWEGIAN</td>
</tr>
<tr>
<td>CROATION</td>
<td>POLISH</td>
</tr>
<tr>
<td>CZECH</td>
<td>PORTUGUESE</td>
</tr>
<tr>
<td>DANISH</td>
<td>ROMANIAN</td>
</tr>
<tr>
<td>DUTCH</td>
<td>RUSSIAN</td>
</tr>
<tr>
<td>ENGLISH</td>
<td>LATIN SERBIAN</td>
</tr>
<tr>
<td>GERMAN</td>
<td>SLOVAK</td>
</tr>
<tr>
<td>GREEK</td>
<td>SWEDISH</td>
</tr>
<tr>
<td>HEBREW</td>
<td>THAI</td>
</tr>
<tr>
<td>HUNGARIAN</td>
<td>TURKISH</td>
</tr>
<tr>
<td>ITALIAN</td>
<td></td>
</tr>
</tbody>
</table>

**CHINESE_VGRAM_LEXER**

The `CHINESE_VGRAM_LEXER` type identifies tokens in Chinese text for creating Text indexes.

**CHINESE_VGRAM_LEXER Attribute**

The `CHINESE_VGRAM_LEXER` has the following attribute:
You can use this lexer if your database uses one of the following character sets:

- AL32UTF8
- ZHS16CGB231280
- ZHS16GBK
- ZHS32GB18030
- ZHT32EUC
- ZHT16BIG5
- ZHT32TRIS
- ZHT16HKSCS
- ZHT16MSWIN950
- UTF8

**CHINESE_LEXER**

The CHINESE_LEXER type identifies tokens in traditional and simplified Chinese text for creating Oracle Text indexes.

This lexer offers the following benefits over the CHINESE_VGRAM_LEXER:

- generates a smaller index
- better query response time
- generates real word tokens resulting in better query precision
- supports stop words

Because the CHINESE_LEXER uses a different algorithm to generate tokens, indexing time is longer than with CHINESE_VGRAM_LEXER.

You can use this lexer if your database character is one of the Chinese or Unicode character sets supported by Oracle.

**CHINESE_LEXER Attribute**

The CHINESE_LEXER has the following attribute:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>mixed_case_ASCII7</td>
<td>Enable mixed-case (upper- and lower-case) searches of ASCII7 text (for example, cat and Cat). Allowable values are <strong>YES</strong> and <strong>NO</strong> (default).</td>
</tr>
</tbody>
</table>
Customizing the Chinese Lexicon
You can modify the existing lexicon (dictionary) used by the Chinese lexer, or create your own Chinese lexicon, with the \texttt{ctxlc} command.

\textbf{See Also: } "Lexical Compiler (ctxlc)" on page \texttt{15-10 in Chapter 15, "Oracle Text Utilities"}

\section*{\texttt{JAPANESE\_VGRAM\_LEXER}}
The \texttt{JAPANESE\_VGRAM\_LEXER} type identifies tokens in Japanese for creating Text indexes. This lexer supports the stem (\$) operator.

\section*{\texttt{JAPANESE\_VGRAM\_LEXER Attributes}}
This lexer has the following attributes:

\begin{table}[H]
\centering
\caption{\texttt{JAPANESE\_VGRAM\_LEXER Attributes}}
\begin{tabular}{ll}
\hline
\textbf{Attribute} & \textbf{Attribute Value} \\
\hline
delimiter & Specify whether to consider certain Japanese blank characters, such as a full-width forward slash or a full-width middle dot. \texttt{ALL} considers these characters, while \texttt{NONE} ignores them. Default is \texttt{NONE}. \\
mixed_case\_ASCII7 & Enable mixed-case (upper- and lower-case) searches of ASCII7 text (for example, \texttt{cat} and \texttt{Cat}). Allowable values are \texttt{YES} and \texttt{NO} (default). \\
\hline
\end{tabular}
\end{table}

\section*{\texttt{JAPANESE\_VGRAM\_LEXER Character Sets}}
You can use this lexer if your database uses one of the following character sets:
\begin{itemize}
\item JA16SJIS
\item JA16EUC
\item UTF8
\item AL32UTF8
\item JA16EUCTILDE
\item JA16EUCYEN
\item JA16SJISTILDE
\item JA16SJISYEN
\end{itemize}

\section*{\texttt{JAPANESE\_LEXER}}
The \texttt{JAPANESE\_LEXER} type identifies tokens in Japanese for creating Text indexes. This lexer supports the stem (\$) operator.

This lexer offers the following benefits over the \texttt{JAPANESE\_VGRAM\_LEXER}:
\begin{itemize}
\item generates a smaller index
\item better query response time
\item generates real word tokens resulting in better query precision
\end{itemize}

Because the \texttt{JAPANESE\_LEXER} uses a new algorithm to generate tokens, indexing time is longer than with \texttt{JAPANESE\_VGRAM\_LEXER}. 

Customizing the Japanese Lexicon
You can modify the existing lexicon (dictionary) used by the Japanese lexer, or create your own Japanese lexicon, with the \texttt{ctxlc} command.

\textbf{See Also:} "Lexical Compiler (ctxlc)" on page 15-10 in Chapter 15, "Oracle Text Utilities"

\textbf{JAPANESE LEXER Attributes}
This lexer has the following attributes:

\begin{table}[h]
\centering
\begin{tabular}{|l|l|}
\hline
\textbf{Attribute} & \textbf{Attribute Value} \\
\hline
 delimiter & Specify \texttt{NONE} or \texttt{ALL} to ignore certain Japanese blank characters, such as a full-width forward slash or a full-width middle dot. Default is \texttt{NONE}. \\
mixed_case.ASCII7 & Enable mixed-case (upper- and lower-case) searches of ASCII text (for example, \texttt{cat} and \texttt{Cat}). Allowable values are \texttt{YES} and \texttt{NO} (default). \\
\hline
\end{tabular}
\end{table}

\textbf{JAPANESE LEXER Character Sets}
The JAPANESE LEXER supports the following character sets:

- JA16SJIS
- JA16EUC
- UTF8
- AL32UTF8
- JA16EUCTILDE
- JA16EUCYEN
- JA16SJISTILDE
- JA16SJISYEN

\textbf{Japanese Lexer Example}
When you specify JAPANESE LEXER for creating text index, the JAPANESE LEXER resolves a sentence into words.

For example, the following compound word (\textit{natural language institute})

\begin{itemize}
\item ['自然言語処理']
\end{itemize}

is indexed as three tokens:

\begin{itemize}
\item ['自然', '言語', '処理']
\end{itemize}
To resolve a sentence into words, the internal dictionary is referenced. When a word cannot be found in the internal dictionary, Oracle Text uses the JAPANESE_VGRAM_LEXER to resolve it.

KOREAN_MORPH_LEXER

The KOREAN_MORPH_LEXER type identifies tokens in Korean text for creating Oracle Text indexes.

Supplied Dictionaries

The KOREAN_MORPH_LEXER uses four dictionaries:

Table 2–29 KOREAN_MORPH_LEXER Dictionaries

<table>
<thead>
<tr>
<th>Dictionary</th>
<th>File</th>
</tr>
</thead>
<tbody>
<tr>
<td>System</td>
<td>$ORACLE_HOME/ctx/data/kolx/drk2sdic.dat</td>
</tr>
<tr>
<td>Grammar</td>
<td>$ORACLE_HOME/ctx/data/kolx/drk2gram.dat</td>
</tr>
<tr>
<td>Stopword</td>
<td>$ORACLE_HOME/ctx/data/kolx/drk2xdic.dat</td>
</tr>
<tr>
<td>User-defined</td>
<td>$ORACLE_HOME/ctx/data/kolx/drk2udic.dat</td>
</tr>
</tbody>
</table>

The grammar, user-defined, and stopword dictionaries should be written using the KSC 5601 or MSWIN949 character sets. You can modify these dictionaries using the defined rules. The system dictionary must not be modified.

You can add unregistered words to the user-defined dictionary file. The rules for specifying new words are in the file.

Supported Character Sets

You can use KOREAN_MORPH_LEXER if your database uses one of the following character sets:

- KO16KSC5601
- KO16MSWIN949
- UTF8
- AL32UTF8

The KOREAN_MORPH_LEXER enables mixed-case searches.

Unicode Support

The KOREAN_MORPH_LEXER supports:

- Words in non-KSC5601 Korean characters defined in Unicode
- Supplementary characters

See Also: For information on supplementary characters, see the Oracle Database Globalization Support Guide

Some Korean documents may have non-KSC5601 characters in them. As the KOREAN_MORPH_LEXER can recognize all possible 11,172 Korean (Hangul) characters, such documents can also be interpreted by using the UTF8 or AL32UTF8 character sets.
Use the AL32UTF8 character set for your database to extract surrogate characters. By default, the KOREAN_MORPH_LEXER extracts all series of surrogate characters in a document as one token for each series.

Limitations on Korean Unicode Support For conversion Hanja to Hangul (Korean), the KOREAN_MORPH_LEXER supports only the 4888 Hanja characters defined in KSC5601.

KOREAN_MORPH_LEXER Attributes

When you use the KOREAN_MORPH_LEXER, you can specify the following attributes:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>verb_adjective</td>
<td>Specify TRUE or FALSE to index verbs, adjectives, and adverbs. Default is FALSE.</td>
</tr>
<tr>
<td>one_char_word</td>
<td>Specify TRUE or FALSE to index one syllable. Default is FALSE.</td>
</tr>
<tr>
<td>number</td>
<td>Specify TRUE or FALSE to index number. Default is FALSE.</td>
</tr>
<tr>
<td>user_dic</td>
<td>Specify TRUE or FALSE to index user dictionary. Default is TRUE.</td>
</tr>
<tr>
<td>stop_dic</td>
<td>Specify TRUE of FALSE to use stop-word dictionary. Default is TRUE. The stop-word dictionary belongs to KOREAN_MORPH_LEXER.</td>
</tr>
<tr>
<td>composite</td>
<td>Specify indexing style of composite noun. Specify COMPOSITE_ONLY to index only composite nouns. Specify NGRAM to index all noun components of a composite noun. Specify COMPONENT_WORD to index single noun components of composite nouns as well as the composite noun itself. Default is COMPONENT_WORD. The following example describes the difference between NGRAM and COMPONENT_WORD.</td>
</tr>
<tr>
<td>morpheme</td>
<td>Specify TRUE or FALSE for morphological analysis. If set to FALSE, tokens are created from the words that are divided by delimiters such as white space in the document. Default is TRUE.</td>
</tr>
<tr>
<td>to_upper</td>
<td>Specify TRUE or FALSE to convert English to uppercase. Default is TRUE.</td>
</tr>
<tr>
<td>hanja</td>
<td>Specify TRUE to index hanja characters. If set to FALSE, hanja characters are converted to hangul characters. Default is FALSE.</td>
</tr>
<tr>
<td>long_word</td>
<td>Specify TRUE to index long words that have more than 16 syllables in Korean. Default is FALSE.</td>
</tr>
<tr>
<td>japanese</td>
<td>Specify TRUE to index Japanese characters in Unicode (only in the 2-byte area). Default is FALSE.</td>
</tr>
<tr>
<td>english</td>
<td>Specify TRUE to index alphanumeric strings. Default is TRUE.</td>
</tr>
</tbody>
</table>

Limitations

Sentence and paragraph sections are not supported with the KOREAN_MORPH_LEXER.

KOREAN_MORPH_LEXER Example: Setting Composite Attribute

Use the composite attribute to control how composite nouns are indexed.

NGRAM Example When you specify NGRAM for the composite attribute, composite nouns are indexed with all possible component tokens. For example, the following composite noun (information processing institute)
is indexed as six tokens:

‘정보’, ‘처리’, ‘학회’, ‘정보처리’

‘처리학회’, ‘정보처리학회’

Specify NGRAM indexing as follows:

begin
ctx_ddl.create_preference('my_lexer','KOREAN_MORPH_LEXER');
ctx_ddl.set_attribute('my_lexer','COMPOSITE','NGRAM');
end

To create the index:

create index koreanx on korean(text) indextype is ctxsys.context
parameters ('lexer my_lexer');

COMPONENT_WORD Example When you specify COMPONENT_WORD for the composite attribute, composite nouns and their components are indexed. For example, the following composite noun (information processing institute)

‘정보처리학회’

is indexed as four tokens:

‘정보처리학회’

‘정보’, ‘처리’, ‘학회’

Specify COMPONENT_WORD indexing as follows:

begin
ctx_ddl.create_preference('my_lexer','KOREAN_MORPH_LEXER');
ctx_ddl.set_attribute('my_lexer','COMPOSITE','COMPONENT_WORD');
end

To create the index:

create index koreanx on korean(text) indextype is ctxsys.context
parameters ('lexer my_lexer');

USER_LEXER

Use USER_LEXER to plug in your own language-specific lexing solution. This enables you to define lexers for languages that are not supported by Oracle Text. It also enables you to define a new lexer for a language that is supported but whose lexer is inappropriate for your application.

The user-defined lexer you register with Oracle Text is composed of two routines that you must supply:
Limitations

The following features are not supported with the USER_LEXER:

- `CTX_DOC.GIST` and `CTX_DOC.THEMES`
- `CTX_QUERY.HFEEDBACK`
- `ABOUT` query operator
- `CTXRULE` index type
- `VGRAM` indexing algorithm

**USER_LEXER Attributes**

USER_LEXER has the following attributes:

<table>
<thead>
<tr>
<th>Table 2–32 USER_LEXER Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attribute</strong></td>
</tr>
<tr>
<td>INDEX_PROCEDURE</td>
</tr>
<tr>
<td>INPUT_TYPE</td>
</tr>
<tr>
<td>QUERY_PROCEDURE</td>
</tr>
</tbody>
</table>

**INDEX_PROCEDURE**

This callback stored procedure is called by Oracle Text as needed to tokenize a document or a stop word found in the stoplist object.

**Requirements**

This procedure can be a PL/SQL stored procedure.

The index owner must have EXECUTE privilege on this stored procedure.

This stored procedure must not be replaced or dropped after the index is created. You can replace or drop this stored procedure after the index is dropped.

**Parameters**

Two different interfaces are supported for the user-defined lexer indexing procedure:

- VARCHAR2 Interface
- CLOB Interface

**Restrictions**

This procedure must not perform any of the following operations:

- Rollback
- Explicitly or implicitly commit the current transaction
- Enter any other transaction control statement
- Alter the session language or territory

The child elements of the root element tokens of the XML document returned must be in the same order as the tokens occur in the document or stop word being tokenized.

The behavior of this stored procedure must be deterministic with respect to all parameters.

**INPUT_TYPE**

Two different interfaces are supported for the User-defined lexer indexing procedure. One interface enables the document or stop word and the corresponding tokens encoded as XML to be passed as VARCHAR2 datatype whereas the other interface uses the CLOB datatype. This attribute indicates the interface implemented by the stored procedure specified by the INDEX_PROCEDURE attribute.

**VARCHAR2 Interface**  
BASIC WORDLIST Attributes Table 2–40 describes the interface that enables the document or stop word from stoplist object to be tokenized to be passed as VARCHAR2 from Oracle Text to the stored procedure and for the tokens to be passed as VARCHAR2 as well from the stored procedure back to Oracle Text.

Your user-defined lexer indexing procedure should use this interface when all documents in the column to be indexed are smaller than or equal to 32512 bytes and the tokens can be represented by less than or equal to 32512 bytes. In this case the CLOB interface given in Table 2–34 can also be used, although the VARCHAR2 interface will generally perform faster than the CLOB interface.

This procedure must be defined with the following parameters:
Table 2–33  VARCHAR2 Interface for INDEX_PROCEDURES

<table>
<thead>
<tr>
<th>Parameter Position</th>
<th>Parameter Mode</th>
<th>Parameter Datatype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IN</td>
<td>VARCHAR2</td>
<td>Document or stop word from stoplist object to be tokenized. If the document is larger than 32512 bytes then Oracle Text will report a document level indexing error.</td>
</tr>
<tr>
<td>2</td>
<td>IN OUT</td>
<td>VARCHAR2</td>
<td>Tokens encoded as XML. If the document contains no tokens, then either NULL must be returned or the tokens element in the XML document returned must contain no child elements. Byte length of the data must be less than or equal to 32512. To improve performance, use the NOCOPY hint when declaring this parameter. This passes the data by reference, rather than passing data by value. The XML document returned by this procedure should not include unnecessary whitespace characters (typically used to improve readability). This reduces the size of the XML document which in turn minimizes the transfer time. To improve performance, index_procedure should not validate the XML document with the corresponding XML schema at run-time. Note that this parameter is IN OUT for performance purposes. The stored procedure has no need to use the IN value.</td>
</tr>
<tr>
<td>3</td>
<td>IN</td>
<td>BOOLEAN</td>
<td>Oracle Text sets this parameter to TRUE when Oracle Text needs the character offset and character length of the tokens as found in the document being tokenized. Oracle Text sets this parameter to FALSE when Text is not interested in the character offset and character length of the tokens as found in the document being tokenized. This implies that the XML attributes off and len must not be used.</td>
</tr>
</tbody>
</table>

CLOB Interface  Table 2–34 describes the CLOB interface that enables the document or stop word from stoplist object to be tokenized to be passed as CLOB from Oracle Text to the stored procedure and for the tokens to be passed as CLOB as well from the stored procedure back to Oracle Text.

The user-defined lexer indexing procedure should use this interface when at least one of the documents in the column to be indexed is larger than 32512 bytes or the corresponding tokens are represented by more than 32512 bytes.
The first and second parameters are temporary CLOBs. Avoid assigning these CLOB locators to other locator variables. Assigning the formal parameter CLOB locator to another locator variable causes a new copy of the temporary CLOB to be created resulting in a performance hit.

**QUERY_PROCEDURE**

This callback stored procedure is called by Oracle Text as needed to tokenize words in the query. A space-delimited group of characters (excluding the query operators) in the query will be identified by Oracle Text as a word.

**Requirements**

This procedure can be a PL/SQL stored procedure.

The index owner must have EXECUTE privilege on this stored procedure.

This stored procedure must not be replaced or be dropped after the index is created. You can replace or drop this stored procedure after the index is dropped.

**Restrictions**

This procedure must not perform any of the following operations:

- Rollback
- Explicitly or implicitly commit the current transaction
- Enter any other transaction control statement
- Alter the session language or territory

The child elements of the root element tokens of the XML document returned must be in the same order as the tokens occur in the query word being tokenized.

The behavior of this stored procedure must be deterministic with respect to all parameters.

---

**Table 2–34 CLOB Interface for INDEX_PROCEDURE**

<table>
<thead>
<tr>
<th>Parameter Position</th>
<th>Parameter Mode</th>
<th>Parameter Datatype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IN</td>
<td>CLOB</td>
<td>Document or stop word from stoplist object to be tokenized.</td>
</tr>
<tr>
<td>2</td>
<td>IN OUT</td>
<td>CLOB</td>
<td>Tokens encoded as XML.</td>
</tr>
<tr>
<td>3</td>
<td>IN</td>
<td>BOOLEAN</td>
<td>If the document contains no tokens, then either NULL must be returned or the tokens element in the XML document returned must contain no child elements. To improve performance, use the <strong>NOCOPY</strong> hint when declaring this parameter. This passes the data by reference, rather than passing data by value. The XML document returned by this procedure should not include unnecessary whitespace characters (typically used to improve readability). This reduces the size of the XML document which in turn minimizes the transfer time. To improve performance, index_procedure should not validate the XML document with the corresponding XML schema at run-time. Note that this parameter is <strong>IN OUT</strong> for performance purposes. The stored procedure has no need to use the <strong>IN</strong> value. The <strong>IN</strong> value will always be a truncated CLOB.</td>
</tr>
</tbody>
</table>

Table 2–34 CLOB Interface for INDEX_PROCEDURE

[Note: This table provides the interface details for the stored procedure as described in Oracle Text Indexing Elements, page 2-59.]
Lexer Types

Parameters Table 2–35 describes the interface for the user-defined lexer query procedure:

Table 2–35 User-defined Lexer Query Procedure XML Schema Attributes

<table>
<thead>
<tr>
<th>Parameter Position</th>
<th>Parameter Mode</th>
<th>Parameter Datatype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IN</td>
<td>VARCHAR2</td>
<td>Query word to be tokenized.</td>
</tr>
<tr>
<td>2</td>
<td>IN</td>
<td>CTX_ULEXER.WILDCARD_TAB</td>
<td>Character offsets of wildcard characters (% and _) in the query word. If the query word passed in by Oracle Text does not contain any wildcard characters then this index-by table will be empty. The wildcard characters in the query word must be preserved in the tokens returned in order for the wildcard query feature to work properly. The character offset is 0 (zero) based. Offset information follows USC-2 codepoint semantics.</td>
</tr>
<tr>
<td>3</td>
<td>IN OUT</td>
<td>VARCHAR2</td>
<td>Tokens encoded as XML. If the query word contains no tokens then either NULL must be returned or the tokens element in the XML document returned must contain no child elements. The length of the data must be less-than or equal to 32512 bytes.</td>
</tr>
</tbody>
</table>

Encoding Tokens as XML

The sequence of tokens returned by your stored procedure must be represented as an XML 1.0 document. The XML document must be valid with respect to the XML Schemas given in the following sections.

- XML Schema for No-Location, User-defined Indexing Procedure
- XML Schema for User-defined Indexing Procedure with Location
- XML Schema for User-defined Lexer Query Procedure

Limitations To boost performance of this feature, the XML parser in Oracle Text will not perform validation and will not be a full-featured XML compliant parser. This implies that only minimal XML features will be supported. The following XML features are not supported:

- Document Type Declaration (for example, \<!DOCTYPE [ . . . ]> ) and therefore entity declarations. Only the following built-in entities can be referenced: lt, gt, amp, quot, and apos.
- CDATA sections.
- Comments.
- Processing Instructions.
- XML declaration (for example, \<?xml version="1.0" ...?>).
- Namespaces.
- Use of elements and attributes other than those defined by the corresponding XML Schema.
- Character references (for example &#{x09F};).
- xml:space attribute.
- xml:lang attribute

**XML Schema for No-Location, User-defined Indexing Procedure**

This section describes additional constraints imposed on the XML document returned by the user-defined lexer indexing procedure when the third parameter is FALSE. The XML document returned must be valid with respect to the following XML Schema:

```xml
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="tokens">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:choice minOccurs="0" maxOccurs="unbounded">
          <xsd:element name="eos" type="EmptyTokenType"/>
          <xsd:element name="eop" type="EmptyTokenType"/>
          <xsd:element name="num" type="xsd:token"/>
          <xsd:element ref="IndexCompositeGroup"/>
        </xsd:choice>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>

  <!-- Enforce constraint that compMem element must be preceeded by word element or compMem element for indexing -->
  <xsd:group name="IndexCompositeGroup">
    <xsd:sequence>
      <xsd:element name="word" type="xsd:token"/>
      <xsd:element name="compMem" type="xsd:token" minOccurs="0" maxOccurs="unbounded"/>
    </xsd:sequence>
  </xsd:group>

  <!-- EmptyTokenType defines an empty element without attributes -->
  <xsd:complexType name="EmptyTokenType"/>
</xsd:schema>
```

Here are some of the constraints imposed by this XML Schema:

- The root element is tokens. This is mandatory. It has no attributes.
- The root element can have zero or more child elements. The child elements can be one of the following elements: eos, eop, num, word, and compMem. Each of these represent a specific type of token.
- The compMem element must be preceded by a word element or a compMem element.
- The eos and eop elements have no attributes and must be empty elements.
- The num, word, and compMem elements have no attributes. Oracle Text will normalize the content of these elements as follows: convert whitespace characters to space characters, collapse adjacent space characters to a single space character, remove leading and trailing spaces, perform entity reference replacement, and truncate to 64 bytes.
Table 2–36 describes the element names defined in the preceding XML Schema.

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>word</td>
<td>This element represents a simple word token. The content of the element is the word itself. Oracle Text does the work of identifying this token as being a stop word or non-stop word and processing it appropriately.</td>
</tr>
<tr>
<td>num</td>
<td>This element represents an arithmetic number token. The content of the element is the arithmetic number itself. Oracle Text treats this token as a stop word if the stoplist preference has NUMBERS added as the stopclass. Otherwise this token is treated the same way as the word token. Supporting this token type is optional. Without support for this token type, adding the NUMBERS stopclass will have no effect.</td>
</tr>
<tr>
<td>eos</td>
<td>This element represents end-of-sentence token. Oracle Text uses this information so that it can support WITHIN SENTENCE queries. Supporting this token type is optional. Without support for this token type, queries against the SENTENCE section will not work as expected.</td>
</tr>
<tr>
<td>eop</td>
<td>This element represents end-of-paragraph token. Oracle Text uses this information so that it can support WITHIN PARAGRAPH queries. Supporting this token type is optional. Without support for this token type, queries against the PARAGRAPH section will not work as expected.</td>
</tr>
<tr>
<td>compMem</td>
<td>Same as the word element, except that the implicit word offset is the same as the previous word token. Support for this token type is optional.</td>
</tr>
</tbody>
</table>

Example Document: Vom Nordhauptbahnhof und aus der Innenstadt zum Messegelände.

Tokens:
```xml
<tokens>
  <word> VOM </word>
  <word> NORDHAUPTBAHNHOF </word>
  <compMem>NORD</compMem>
  <compMem>HAUPT </compMem>
  <compMem>BAHNHOF </compMem>
  <compMem>HAUPTBAHNHOF </compMem>
  <word> UND </word>
  <word> AUS </word>
  <word> DER </word>
  <word> INNENSTADT </word>
  <word> ZUM </word>
  <word> MESSEGLÄNDE </word>
  <eos/>
</tokens>
```

Example Document: Oracle Database 11g Release 1

Tokens:
```xml
<tokens>
  <word> ORACLE11G </word>
  <word> RELEASE </word>
  <num> 1 </num>
</tokens>
```

Example Document: WHERE salary<25000.00 AND job = 'F&B Manager'
Lexer Types

Tokens:
<tokens>
<word> WHERE </word>
<word> salary&lt;2500.00 </word>
<word> AND </word>
<word> job </word>
<word> F&amp;B </word>
<word> Manager </word>
</tokens>

XML Schema for User-defined Indexing Procedure with Location
This section describes additional constraints imposed on the XML document returned by the user-defined lexer indexing procedure when the third parameter is TRUE. The XML document returned must be valid according to the following XML schema:

```xml
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="tokens">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:choice minOccurs="0" maxOccurs="unbounded">
          <xsd:element name="eos" type="EmptyTokenType"/>
          <xsd:element name="eop" type="EmptyTokenType"/>
          <xsd:element name="num" type="DocServiceTokenType"/>
          <xsd:group ref="DocServiceCompositeGroup"/>
        </xsd:choice>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>

  <!-- Enforce constraint that compMem element must be preceded by word element or compMem element for document service -->
  <xsd:group name="DocServiceCompositeGroup">
    <xsd:sequence>
      <xsd:element name="word" type="DocServiceTokenType"/>
      <xsd:element name="compMem" type="DocServiceTokenType" minOccurs="0" maxOccurs="unbounded"/>
    </xsd:sequence>
  </xsd:group>

  <!-- EmptyTokenType defines an empty element without attributes -->
  <xsd:complexType name="EmptyTokenType"/>

  <!-- DocServiceTokenType defines an element with content and mandatory attributes -->
  <xsd:complexType name="DocServiceTokenType">
    <xsd:simpleContent>
      <xsd:extension base="xsd:token">
        <xsd:attribute name="off" type="OffsetType" use="required"/>
        <xsd:attribute name="len" type="xsd:unsignedShort" use="required"/>
      </xsd:extension>
    </xsd:simpleContent>
  </xsd:complexType>

  <xsd:simpleType name="OffsetType">
    <xsd:restriction base="xsd:unsignedInt">
      <!-- empty -->
    </xsd:restriction>
  </xsd:simpleType>
</xsd:schema>
```
Some of the constraints imposed by this XML Schema are as follows:

- The root element is tokens. This is mandatory. It has no attributes.
- The root element can have zero or more child elements. The child elements can be one of the following elements: eos, eop, num, word, and compMem. Each of these represent a specific type of token.
- The compMem element must be preceded by a word element or a compMem element.
- The eos and eop elements have no attributes and must be empty elements.
- The num, word, and compMem elements have two mandatory attributes: \( off \) and \( len \). Oracle Text will normalize the content of these elements as follows: convert whitespace characters to space characters, collapse adjacent space characters to a single space character, remove leading and trailing spaces, perform entity reference replacement, and truncate to 64 bytes.
- The \( off \) attribute value must be an integer between 0 and 2147483647 inclusive.
- The \( len \) attribute value must be an integer between 0 and 65535 inclusive.

Table 2–36 describes the element types defined in the preceding XML Schema. Table 2–37 describes the attributes defined in the preceding XML Schema.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>off</td>
<td>This attribute represents the character offset of the token as it appears in the document being tokenized. The offset is with respect to the character document passed to the user-defined lexer indexing procedure, not the document fetched by the datastore. The document fetched by the datastore may be pre-processed by the filter object or the section group object, or both, before being passed to the user-defined lexer indexing procedure. The offset of the first character in the document being tokenized is 0 (zero). Offset information follows USC-2 codepoint semantics.</td>
</tr>
<tr>
<td>len</td>
<td>This attribute represents the character length (same semantics as SQL function <code>LENGTH</code>) of the token as it appears in the document being tokenized. The length is with respect to the character document passed to the user-defined lexer indexing procedure, not the document fetched by the datastore. The document fetched by the datastore may be pre-processed by the filter object or the section group object before being passed to the user-defined lexer indexing procedure. Length information follows USC-2 codepoint semantics.</td>
</tr>
</tbody>
</table>

Sum of \( off \) attribute value and \( len \) attribute value must be less than or equal to the total number of characters in the document being tokenized. This is to ensure that the document offset and characters being referenced are within the document boundary.

**Example** Document: User-defined Lexer.
Tokens:

```xml
<tokens>
  <word off="0" len="4"> USE </word>
  <word off="5" len="7"> DEF </word>
  <word off="13" len="5"> LEX </word>
  <eos/>
</tokens>
```

XML Schema for User-defined Lexer Query Procedure

This section describes additional constraints imposed on the XML document returned by the user-defined lexer query procedure. The XML document returned must be valid with respect to the following XML Schema:

```xml
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="tokens">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:choice minOccurs="0" maxOccurs="unbounded">
          <xsd:element name="num" type="QueryTokenType"/>
          <xsd:group ref="QueryCompositeGroup"/>
        </xsd:choice>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

```xml
<!-- Enforce constraint that compMem element must be preceded by word element or compMem element for query -->
<xsd:group name="QueryCompositeGroup">
  <xsd:sequence>
    <xsd:element name="word" type="QueryTokenType"/>
    <xsd:element name="compMem" type="QueryTokenType" minOccurs="0" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:group>
```

```xml
<!-- QueryTokenType defines an element with content and with an optional attribute -->
<xsd:complexType name="QueryTokenType">
  <xsd:simpleContent>
    <xsd:extension base="xsd:token">
      <xsd:attribute name="wildcard" type="WildcardType" use="optional"/>
    </xsd:extension>
  </xsd:simpleContent>
</xsd:complexType>
```

```xml
<xsd:simpleType name="WildcardType">
  <xsd:restriction base="WildcardBaseType">
    <xsd:minLength value="1"/>
    <xsd:maxLength value="64"/>
  </xsd:restriction>
</xsd:simpleType>
```

```xml
<xsd:simpleType name="WildcardBaseType">
  <xsd:list/>
</xsd:simpleType>
```
Here are some of the constraints imposed by this XML Schema:

- The root element is tokens. This is mandatory. It has no attributes.
- The root element can have zero or more child elements. The child elements can be one of the following elements: num and word. Each of these represent a specific type of token.
- The compMem element must be preceded by a word element or a compMem element.

The purpose of compMem is to enable USER_LEXER queries to return multiple forms for a single query. For example, if a user-defined lexer indexes the word bank as BANK(FINANCIAL) and BANK(RIVER), the query procedure can return the first term as a word and the second as a compMem element:

```
<tokens>
  <word>BANK(RIVER)</word>
  <compMem>BANK(FINANCIAL)</compMem>
</tokens>
```

See Table 2–38, "User-defined Lexer Query Procedure XML Schema Attributes" on page 2-66 for more on the compMem element.

- The num and word elements have a single optional attribute: wildcard. Oracle Text will normalize the content of these elements as follows: convert whitespace characters to space characters, collapse adjacent space characters to a single space character, remove leading and trailing spaces, perform entity reference replacement, and truncate to 64 bytes.
- The wildcard attribute value is a white-space separated list of integers. The minimum number of integers is 1 and the maximum number of integers is 64. The value of the integers must be between 0 and 378 inclusive. The intriguers in the list can be in any order.

Table 2–36 describes the element types defined in the preceding XML Schema.
Table 2–38 describes the attribute defined in the preceding XML Schema.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>compMem</td>
<td>Same as the word element, but its implicit word offset is the same as the previous word token. Oracle Text will equate this token with the previous word token and with subsequent compMem tokens using the query EQUIV operator.</td>
</tr>
</tbody>
</table>
Example Query word: pseudo-%morph%

Tokens:
<tokens>
  <word> PSEUDO </word>
  <word wildcard="1 7"> %MORPH% </word>
</tokens>

Example Query word: <%
Tokens:
<tokens>
  <word wildcard="5"> &lt;%= </word>
</tokens>

WORLD_LEXER

Use the WORLD_LEXER to index text columns that contain documents of different languages. For example, use this lexer to index a text column that stores English, Japanese, and German documents.

WORLD_LEXER differs from MULTI_LEXER in that WORLD_LEXER automatically detects the language(s) of a document. Unlike MULTI_LEXER, WORLD_LEXER does not require you to have a language column in your base table nor to specify the language column when you create the index. Moreover, it is not necessary to use sub-lexers, as with MULTI_LEXER. (See "MULTI_LEXER" on page 2-48.)

WORLD_LEXER supports all database character sets, and for languages whose character sets are Unicode-based, it supports the Unicode 5.0 standard. For a list of languages that WORLD_LEXER can work with, see "World Lexer Features" on page D-4.

WORLD_LEXER Attribute

The WORLD_VGRAM_LEXER has the following attribute:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>mixed_case</td>
<td>Enable mixed-case (upper- and lower-case) searches of text (for example, cat and Cat). Allowable values are YES and NO (default).</td>
</tr>
</tbody>
</table>
**WORLD_LEXER Example**

Here is an example of creating an index using `WORLD_LEXER`.

```sql
exec ctx_ddl.create_preference('MYLEXER', 'world_lexer');
create index doc_idx on doc(data)
  indextype is CONTEXT
  parameters ('lexer MYLEXER
    stoplist CTXSYS.EMPTY_STOPLIST');
```

**Wordlist Type**

Use the wordlist preference to enable the query options such as stemming, fuzzy matching for your language. You can also use the wordlist preference to enable substring and prefix indexing, which improves performance for wildcard queries with `CONTAINS` and `CATSEARCH`.

To create a wordlist preference, you must use `BASIC_WORDLIST`, which is the only type available.

**BASIC_WORDLIST**

Use `BASIC_WORDLIST` type to enable stemming and fuzzy matching or to create prefix indexes with Text indexes.

**See Also:** Chapter 3, "Oracle Text `CONTAINS` Query Operators"

`BASIC_WORDLIST` has the following attributes:

**Table 2–40 BASIC_WORDLIST Attributes**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>stemmer</td>
<td>Specifying which language stemmer to use. You can specify one of the following stemmers: NULL (no stemming) ENGLISH (English inflectional) DERIVATIONAL (English derivational) DUTCH FRENCH GERMAN ITALIAN SPANISH AUTO (Automatic language-detection for stemming, derived from the database session language. If the database session language is AMERICAN or ENGLISH, then the ENGLISH stemmer is used. Does not auto-detect JAPANESE.) JAPANESE</td>
</tr>
</tbody>
</table>
### Table 2-40 (Cont.) BASIC_WORDLIST Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>fuzzy_match</td>
<td>Specify which fuzzy matching cluster to use. You can specify one of the following types: AUTO (Automatic language detection for stemming.) CHINESE_VGRAM DUTCH ENGLISH FRENCH GENERIC GERMAN ITALIAN JAPANESE_VGRAM KOREAN OCR SPANISH</td>
</tr>
<tr>
<td>fuzzy_score</td>
<td>Specify a default lower limit of fuzzy score. Specify a number between 0 and 80. Text with scores below this number is not returned. Default is 60.</td>
</tr>
<tr>
<td>fuzzy_numresults</td>
<td>Specify the maximum number of fuzzy expansions. Use a number between 0 and 5,000. Default is 100.</td>
</tr>
<tr>
<td>substring_index</td>
<td>Specify TRUE for Oracle Text to create a substring index. A substring index improves left-truncated and double-truncated wildcard queries such as %ing or %benz%. Default is FALSE. To create a Text index with a wordlist that has substring_index set to TRUE, the user creating the index needs the CREATE TRIGGER system privilege.</td>
</tr>
<tr>
<td>prefix_index</td>
<td>Specify TRUE to enable prefix indexing. Prefix indexing improves performance for right truncated wildcard searches such as TO%. Default is FALSE.</td>
</tr>
<tr>
<td>prefix_min_length</td>
<td>Specify the minimum length of indexed prefixes. Default is 1. Length information must follow USC-2 codepoint semantics.</td>
</tr>
<tr>
<td>prefix_max_length</td>
<td>Specify the maximum length of indexed prefixes. Default is 64. Length information must follow USC-2 codepoint semantics.</td>
</tr>
<tr>
<td>wildcard_maxterms</td>
<td>Specify the maximum number of terms in a wildcard expansion. The maximum value is 50000 and the default value is 20000. If you specify a value of 0, then the number of wildcard expansions will be unbounded. Note that when set to 0, the system may run out of memory due to the high number of wildcard expansions.</td>
</tr>
<tr>
<td>ndata_base_letter</td>
<td>Specify whether characters that have diacritical marks are converted to their base form before being stored in the Text index or queried by the NDATA operator. FALSE (default) or TRUE When set to FALSE, no base lettering is used.</td>
</tr>
</tbody>
</table>
stemmer
Specify the stemmer used for word stemming in Text queries. When you do not specify a value for STEMMER, the default is ENGLISH.

Specify AUTO for the system to automatically set the stemming language according to the language setting of the database session. If the database language is AMERICAN or ENGLISH, then the ENGLISH stemmer is automatically used. Otherwise, the stemmer that maps to the database session language is used.

When there is no stemmer for a language, the default is NULL. With the NULL stemmer, the stem operator is ignored in queries.

You can create your own stemming user-dictionary. See "Stemming User-Dictionaries" on page 2-45 for more information.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndata_alternate_spelling</td>
<td>Specify whether to enable alternate spelling for German, Danish, and Swedish. Enabling alternate spelling allows you to index NDATA section data and query using the NDATA operator in alternate form.</td>
</tr>
<tr>
<td></td>
<td>FALSE (default) or TRUE</td>
</tr>
<tr>
<td></td>
<td>When set to FALSE, no alternate spelling is used.</td>
</tr>
<tr>
<td>ndata_thesaurus</td>
<td>Name of the thesaurus used for alternate name expansion.</td>
</tr>
<tr>
<td>ndata_join_particles</td>
<td>A list of colon-separated name particles that can be joined with a name that follows them.</td>
</tr>
</tbody>
</table>

**Note:** The STEMMER attribute of BASIC_WORDLIST preference will be ignored if:

1. INDEX_STEMS attribute of BASIC_LEXER preference is set to BOKMAL, CATALAN, CROATIAN, CZECH, DANISH, FINNISH, GREEK, HEBREW, HUNGARIAN, NYNORSK, POLISH, PORTUGUESE, ROMANIAN, RUSSIAN, SERBIAN, SLOVAK, SLOVENIAN, SWEDISH, ENGLISH_NEW, DERIVATIONAL_NEW, DUTCH_NEW, FRENCH_NEW, GERMAN_NEW, ITALIAN_NEW, or SPANISH_NEW.
   Or
2. INDEX_STEMS attribute of AUTO_LEXER preference is set to YES.
   Or
3. The database session language causes MULTI_LEXER to choose a SUB_LEXER with the same setting as 1 or 2 above.

In these cases, the same stemmer that is used by the BASIC_LEXER or AUTO_LEXER during indexing will be used to determine the stem of the query term during query.

fuzzy_match
Specify which fuzzy matching routines are used for the column. Fuzzy matching is currently supported for English, Japanese, and, to a lesser extent, the Western European languages.
Note:  The fuzzy_match attributes value for Chinese and Korean are dummy attribute values that prevent the English and Japanese fuzzy matching routines from being used on Chinese and Korean text.

The default for fuzzy_match is GENERIC.

Specify AUTO for the system to automatically set the fuzzy matching language according to language setting of the session.

fuzzy_score
Specify a default lower limit of fuzzy score. Specify a number between 0 and 80. Text with scores below this number are not returned. The default is 60.

Fuzzy score is a measure of how close the expanded word is to the query word. The higher the score the better the match. Use this parameter to limit fuzzy expansions to the best matches.

fuzzy_numresults
Specify the maximum number of fuzzy expansions. Use a number between 0 and 5000. The default is 100.

Setting a fuzzy expansion limits the expansion to a specified number of the best matching words.

substring_index
Specify TRUE for Oracle Text to create a substring index. A substring index improves performance for left-truncated or double-truncated wildcard queries such as %ing or %benz%. The default is false.

Substring indexing has the following impact on indexing and disk resources:
- Index creation and DML processing is up to 4 times slower
- Index creation with substring_index enabled requires more rollback segments during index flushes than with substring index off. Oracle recommends that you do either of the following when creating a substring index:
  - Make available double the usual rollback or
  - Decrease the index memory to reduce the size of the index flushes to disk

prefix_index
Specify yes to enable prefix indexing. Prefix indexing improves performance for right truncated wildcard searches such as TO%. Default is NO.

Note:  Enabling prefix indexing increases index size.

Prefix indexing chops up tokens into multiple prefixes to store in the $I table. For example, words TOKEN and TOY are normally indexed as follows in the $I table:

<table>
<thead>
<tr>
<th>Token</th>
<th>Type</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOKEN</td>
<td>0</td>
<td>DOCID 1 POS 1</td>
</tr>
<tr>
<td>TOY</td>
<td>0</td>
<td>DOCID 1 POS 3</td>
</tr>
</tbody>
</table>
With prefix indexing, Oracle Text indexes the prefix substrings of these tokens as follows with a new token type of 6:

<table>
<thead>
<tr>
<th>Token</th>
<th>Type</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOKEN</td>
<td>0</td>
<td>DOCID 1 POS 1</td>
</tr>
<tr>
<td>TOY</td>
<td>0</td>
<td>DOCID 1 POS 3</td>
</tr>
<tr>
<td>T</td>
<td>6</td>
<td>DOCID 1 POS 1 POS 3</td>
</tr>
<tr>
<td>TO</td>
<td>6</td>
<td>DOCID 1 POS 1 POS 3</td>
</tr>
<tr>
<td>TOK</td>
<td>6</td>
<td>DOCID 1 POS 1</td>
</tr>
<tr>
<td>TOKE</td>
<td>6</td>
<td>DOCID 1 POS 1</td>
</tr>
<tr>
<td>TOKEN</td>
<td>6</td>
<td>DOCID 1 POS 1</td>
</tr>
<tr>
<td>TOY</td>
<td>6</td>
<td>DOCID 1 POS 3</td>
</tr>
</tbody>
</table>

Wildcard searches such as TO% are now faster because Oracle Text does no expansion of terms and merging of result sets. To obtain the result, Oracle Text need only examine the (TO,6) row.

**prefix_min_length**
Specify the minimum length of indexed prefixes. Default is 1.

For example, setting `prefix_min_length` to 3 and `prefix_max_length` to 5 indexes all prefixes between 3 and 5 characters long.

**Note:** A wildcard search whose pattern is below the minimum length or above the maximum length is searched using the slower method of equivalence expansion and merging.

**prefix_max_length**
Specify the maximum length of indexed prefixes. Default is 64.

For example, setting `prefix_min_length` to 3 and `prefix_max_length` to 5 indexes all prefixes between 3 and 5 characters long.

**Note:** A wildcard search whose pattern is below the minimum length or above the maximum length is searched using the slower method of equivalence expansion and merging.

**wildcard_maxterms**
Specify the maximum number of terms in a wildcard (%) expansion. Use this parameter to keep wildcard query performance within an acceptable limit. Oracle Text returns an error when the wildcard query expansion exceeds this number.

**ndata_base_letter**
Specify whether characters that have diacritical marks (umlauts, cedillas, acute accents, and so on) are converted to their base form before being stored in the Text index or queried by the NDATA operator. The default is `FALSE` (base-letter conversion disabled). For more information on base-letter conversions, see "Base-Letter Conversion" on page 16-2.
**ndata_alternate_spelling**
Specify whether to enable alternate spelling for German, Danish, and Swedish. Enabling alternate spelling allows you to index NDATA section data and query using the NDATA operator in alternate form.

When `ndata_base_letter` is enabled at the same time as `ndata_alternate_spelling`, NDATA section data is serially transformed first by alternate spelling and then by base lettering. For more information about the alternate spelling conventions Oracle Text uses, see "Alternate Spelling" on page 16-2.

**ndata_thesaurus**
Specify a name of the thesaurus used for alternate name expansion. The indexing engine expands names in documents using synonym rings in the thesaurus. A user should make use of homographic disambiguating feature of the thesaurus to distinguish common nicknames.

An example is:

```
Albert
  SYN A1
  SYN Bert
Alfred
  SYN A1
  SYN Fred
```

A simple definition such as the above will put Albert, Alfred, Al, Bert, and Fred into the same synonym ring. This will cause an unexpected expansion such that the expansion of Bert includes Fred. To prevent this, you can use homographic disambiguation as in:

```
Albert
  SYN A1 (Albert)
  SYN Bert (Albert)
Alfred
  SYN A1 (Alfred)
  SYN Fred (Alfred)
```

This forms two synonym rings, Albert-Al-Bert and Alfred-Al-Fred. Thus, the expansion of Bert no longer includes Fred. A more detailed example is:

```
begin
  ctx_ddl.create_preference('NDATA_PREF', 'BASIC_WORDLIST');
  ctx_ddl.set_attribute('NDATA_PREF', 'NDATA_ALTERNATE_SPELLING', 'FALSE');
  ctx_ddl.set_attribute('NDATA_PREF', 'NDATA_BASE_LETTER', 'TRUE');
  ctx_ddl.set_attribute('NDATA_PREF', 'NDATA_THESAURUS', 'NICKNAMES');
end;
```

**Note:** A sample thesaurus for names can be found in the `$ORACLE_HOME/ctx/sample/thes` directory. This file is `dr0thsnames.txt`.

**ndata_join_particles**
Specify a list of colon-separated name particles that can be joined with a name that follows them. A name particle, such as da, is written separately from or joined with its following name like da Vinci or daVinci. The indexing engine generates index data for both separated and join versions of a name when it finds a name particle specified in this preference. The same happens in the query processing for better recall.
**BASIC_WORDLIST Example**

The following example shows the use of the **BASIC_WORDLIST** type.

**Enabling Fuzzy Matching and Stemming**
The following example enables stemming and fuzzy matching for English. The preference **STEM_FUZZY_PREF** sets the number of expansions to the maximum allowed. This preference also instructs the system to create a substring index to improve the performance of double-truncated searches.

```
begin
    ctx_ddl.create_preference('STEM_FUZZY_PREF', 'BASIC_WORDLIST');
    ctx_ddl.set_attribute('STEM_FUZZY_PREF','FUZZY_MATCH','ENGLISH');
    ctx_ddl.set_attribute('STEM_FUZZY_PREF','FUZZY_SCORE','0');
    ctx_ddl.set_attribute('STEM_FUZZY_PREF','FUZZY_NUMRESULTS','5000');
    ctx_ddl.set_attribute('STEM_FUZZY_PREF','SUBSTRING_INDEX','TRUE');
    ctx_ddl.set_attribute('STEM_FUZZY_PREF','STEMMER','ENGLISH');
end;
```

To create the index in SQL, enter the following statement:

```
create index fuzzy_stem_subst_idx on mytable ( docs )
    indextype is ctxsys.context parameters ('Wordlist STEM_FUZZY_PREF');
```

**Enabling Sub-string and Prefix Indexing**
The following example sets the wordlist preference for prefix and sub-string indexing. For prefix indexing, it specifies that Oracle Text create token prefixes between 3 and 4 characters long:

```
begin
    ctx_ddl.create_preference('mywordlist', 'BASIC_WORDLIST');
    ctx_ddl.set_attribute('mywordlist','PREFIX_INDEX','TRUE');
    ctx_ddl.set_attribute('mywordlist','PREFIX_MIN_LENGTH',3);
    ctx_ddl.set_attribute('mywordlist','PREFIX_MAX_LENGTH', 4);
    ctx_ddl.set_attribute('mywordlist','SUBSTRING_INDEX', 'YES');
end
```

**Setting Wildcard Expansion Limit**

Use the wildcard_maxterms attribute to set the maximum allowed terms in a wildcard expansion.

```
--- create a sample table
drop table quick ;
create table quick
    ( quick_id number primary key,
        text      varchar(80) )
;
--- insert a row with 10 expansions for 'tire%'
insert into quick ( quick_id, text )
    values ( 1, 'tire tirea tireb tired tiree tiref tireg tireh tirei tirej');
commit;

--- create an index using wildcard_maxterms=100
begin
    Ctx_Ddl.Create_Preference('wildcard_pref', 'BASIC_WORDLIST');
    ctx_ddl.set_attribute('wildcard_pref', 'wildcard_maxterms', 100) ;
end;
```
/ 
create index wildcard_idx on quick(text) 
  indextype is ctxsys.context 
  parameters ('Wordlist wildcard_pref') ;

--- query on 'tire%' - should work fine 
select quick_id from quick 
  where contains ( text, 'tire%' ) > 0;

--- now re-create the index with wildcard_maxterms=5 
drop index wildcard_idx ;

begin 
 Ctx_Ddl.Drop_Preference('wildcard_pref'); 
  Ctx_Ddl.Create_Preference('wildcard_pref', 'BASIC_WORDLIST'); 
  ctx_ddl.set_attribute('wildcard_pref', 'wildcard_maxterms', 5) ;
end;
/

create index wildcard_idx on quick(text) 
  indextype is ctxsys.context 
  parameters ('Wordlist wildcard_pref') ;

--- query on 'tire%' gives "wildcard query expansion resulted in too many terms" 
select quick_id from quick 
  where contains ( text, 'tire%' ) > 0;

Storage Types

Use the storage preference to specify tablespace and creation parameters for tables 
associated with a Text index. The system provides a single storage type called BASIC_ 
STORAGE:

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASIC_STORAGE</td>
<td>Indexing type used to specify the tablespace and creation parameters for the database tables and indexes that constitute a Text index.</td>
</tr>
</tbody>
</table>

BASIC_STORAGE

The BASIC_STORAGE type specifies the tablespace and creation parameters for the 
database tables and indexes that constitute a Text index.

The clause you specify is added to the internal CREATE TABLE (CREATE INDEX for the 
i_index_clause) statement at index creation. You can specify most allowable 
clauses, such as storage, LOB storage, or partitioning. However, you cannot specify an 
index organized table clause.

See Also: For more information about how to specify CREATE TABLE and CREATE INDEX statements, see Oracle Database SQL 
Language Reference.

BASIC_STORAGE has the following attributes:
Table 2–42  BASIC_STORAGE Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>i_index_clause</td>
<td>Parameter clause for dr$indexname$X index creation. Specify storage and tablespace clauses to add to the end of the internal CREATE INDEX statement. The default clause is: ‘COMPRESS 2’ which instructs Oracle Text to compress this index table. If you choose to override the default, Oracle recommends including COMPRESS 2 in your parameter clause to compress this table, because such compression saves disk space and helps query performance.</td>
</tr>
<tr>
<td>i_rowid_index_clause</td>
<td>Parameter clause to specify the storage clause for the $R index on dr$rowid column of the $I table. Specify storage and tablespace clauses to add to the end of the internal CREATE INDEX statement. This clause is only used by the CTXCAT index type.</td>
</tr>
<tr>
<td>i_table_clause</td>
<td>Parameter clause for dr$indexname$I table creation. Specify storage and tablespace clauses to add to the end of the internal CREATE TABLE statement. The I table is the index data table. <strong>Note:</strong> Oracle strongly recommends that you do not specify &quot;disable storage in row&quot; for $I LOBs, as this will greatly degrade the query performance.</td>
</tr>
<tr>
<td>k_table_clause</td>
<td>Parameter clause for dr$indexname$K table creation. Specify storage and tablespace clauses to add to the end of the internal CREATE TABLE statement. The K table is the keymap table.</td>
</tr>
<tr>
<td>r_table_clause</td>
<td>Parameter clause for dr$indexname$R table creation. Specify storage and tablespace clauses to add to the end of the internal CREATE TABLE statement. The R table is the rowid table. The default clause is: ‘LOB(DATA) STORE AS (CACHE)’. If you modify this attribute, always include this clause for good performance.</td>
</tr>
<tr>
<td>n_table_clause</td>
<td>Parameter clause for dr$indexname$N table creation. Specify storage and tablespace clauses to add to the end of the internal CREATE TABLE statement. The N table is the negative list table.</td>
</tr>
<tr>
<td>p_table_clause</td>
<td>Parameter clause for the substring index if you have enabled SUBSTRING_INDEX in the BASIC_WORDLIST. Specify storage and tablespace clauses to add to the end of the internal CREATE INDEX statement. The P table is an index-organized table so the storage clause you specify must be appropriate to this type of table.</td>
</tr>
</tbody>
</table>
By default, BASIC_STORAGE attributes are not set. In such cases, the Text index tables are created in the index owner's default tablespace. Consider the following statement, entered by user IUSER, with no BASIC_STORAGE attributes set:

```
create index IOWNER.idx on TOWNER.tab(b) indextype is ctxsys.context;
```

In this example, the text index is created in IOWNER's default tablespace.

### Storage Examples

The following examples specify that the index tables are to be created in the foo tablespace with an initial extent of 1K:

```sql
begin
ctx_ddl.create_preference('mystore', 'BASIC_STORAGE');
ctx_ddl.set_attribute('mystore', 'I_TABLE_CLAUSE',
  'tablespace foo storage (initial 1K)');
ctx_ddl.set_attribute('mystore', 'K_TABLE_CLAUSE',
  'tablespace foo storage (initial 1K)');
ctx_ddl.set_attribute('mystore', 'R_TABLE_CLAUSE',
  'tablespace users storage (initial 1K) lob (data) store as (disable storage in row cache)');
ctx_ddl.set_attribute('mystore', 'N_TABLE_CLAUSE',
  'tablespace foo storage (initial 1K)');
ctx_ddl.set_attribute('mystore', 'I_INDEX_CLAUSE',
  'tablespace foo storage (initial 1K) compress 2');
ctx_ddl.set_attribute('mystore', 'P_TABLE_CLAUSE',
  'tablespace foo storage (initial 1K)');
ctx_ddl.set_attribute('mystore', 'S_TABLE_CLAUSE',
  'tablespace foo storage (initial 1K)');
end;
```

### Section Group Types

To enter WITHIN queries on document sections, you must create a section group before you define your sections. Specify your section group in the parameter clause of CREATE INDEX.

To create a section group, you can specify one of the following group types with the CTX_DDL.CREATESECTIONGROUP procedure:

---

**Table 2–42 (Cont.) BASIC_STORAGE Attributes**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>s_table_clause</td>
<td>Parameter clause for dr$indexname$S table creation*. Specify storage and tablespace clauses to add to the end of the internal CREATE TABLE statement. The default clause is nocompres.</td>
</tr>
</tbody>
</table>

* For performance reasons, $S table must be created on a tablespace with db block size >= 4K without overflow segment and without a PCTTHRESHOLD clause. If $S is created on a tablespace with db block size < 4K, or is created with an overflow segment or with PCTTHRESHOLD clause, then appropriate errors will be raised during CREATE INDEX.

The S table is the table that stores SDATA section values.

If this clause is specified for a storage preference in an index without SDATA, then it will have no effect on the index, and index creation will still succeed.
Table 2–43  Section Group Types

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL_SECTION_GROUP</td>
<td>Use this group type when you define no sections or when you define only SENTENCE or PARAGRAPH sections. This is the default.</td>
</tr>
<tr>
<td>BASIC_SECTION_GROUP</td>
<td>Use this group type for defining sections where the start and end tags are of the form <code>&lt;a&gt;</code> and <code>&lt;/a&gt;</code>. Note: This group type does not support input such as unbalanced parentheses, comments tags, and attributes. Use HTML_SECTION_GROUP for this type of input.</td>
</tr>
<tr>
<td>HTML_SECTION_GROUP</td>
<td>Use this group type for indexing HTML documents and for defining sections in HTML documents.</td>
</tr>
<tr>
<td>XML_SECTION_GROUP</td>
<td>Use this group type for indexing XML documents and for defining sections in XML documents. All sections to be indexed must be manually defined for this group.</td>
</tr>
</tbody>
</table>
| AUTO_SECTION_GROUP | Use this group type to automatically create a zone section for each start-tag/end-tag pair in an XML document. The section names derived from XML tags are case sensitive as in XML. Attribute sections are created automatically for XML tags that have attributes. Attribute sections are named in the form tag@attribute. Special sections can be added to AUTO_SECTION_GROUP for WITHIN SENTENCE and WITHIN PARAGRAPH searches. Once a sentence or paragraph section is added to the AUTO_SECTION_GROUP, sections with corresponding tag names 'sentence' or 'paragraph' (case insensitive) are treated as stop sections. Stop sections, empty tags, processing instructions, and comments are not indexed. The following limitations apply to automatic section groups:  
- You cannot add zone, field, or special sections to an automatic section group.  
- You can define a stop section that applies only to one particular type; that is, if you have two different XML DTDs, both of which use a tag called FOO, you can define (TYPE1)FOO to be stopped, but (TYPE2)FOO to not be stopped.  
- The length of the indexed tags, including prefix and namespace, cannot exceed 64 bytes. Tags longer than this are not indexed. |
| PATH_SECTION_GROUP | Use this group type to index XML documents. Behaves like the AUTO_SECTION_GROUP. The difference is that with this section group you can do path searching with the INPATH and HASPATH operators. Queries are also case-sensitive for tag and attribute names. Stop sections are not allowed. |
| NEWS_SECTION_GROUP | Use this group for defining sections in newsgroup formatted documents according to RFC 1036.                                               |

Section Group Examples

This example shows the use of section groups in both HTML and XML documents.
Creating Section Groups in HTML Documents
The following statement creates a section group called `htmgroup` with the HTML group type.

begin
  ctx_ddl.create_section_group('htmgroup', 'HTML_SECTION_GROUP');
end;

You can optionally add sections to this group using the procedures in the CTX_DDL package, such as `CTX_DDL.ADD_SPECIAL_SECTION` or `CTX_DDL.ADD_ZONE_SECTION`. To index your documents, enter a statement such as:

create index myindex on docs(htmlfile) indextype is ctxsys.context
  parameters('filter ctxsys.null_filter section group htmgroup');

See Also: For more information on section groups, see Chapter 7, "CTX_DDL Package"

Creating Sections Groups in XML Documents
The following statement creates a section group called `xmlgroup` with the XML_SECTION_GROUP group type.

begin
  ctx_ddl.create_section_group('xmlgroup', 'XML_SECTION_GROUP');
end;

You can optionally add sections to this group using the procedures in the CTX_DDL package, such as `CTX_DDL.ADD_ATTR_SECTION` or `CTX_DDL.ADD_STOP_SECTION`. To index your documents, enter a statement such as:

create index myindex on docs(htmlfile) indextype is ctxsys.context
  parameters('filter ctxsys.null_filter section group xmlgroup');

See Also: For more information on section groups, see Chapter 7, "CTX_DDL Package"

Automatic Sectioning in XML Documents
The following statement creates a section group called `auto` with the AUTO_SECTION_GROUP group type. This section group automatically creates sections from tags in XML documents.

begin
  ctx_ddl.create_section_group('auto', 'AUTO_SECTION_GROUP');
end;

CREATE INDEX myindex on docs(htmlfile) INDEXTYPE IS ctxsys.context
  PARAMETERS('filter ctxsys.null_filter section group auto');

Classifier Types
This section describes the classifier types used to create a preference for CTX_CLS.TRAIN and CTXRULE index creation. The following two classifier types are supported:

- RULE_CLASSIFIER
- SVM_CLASSIFIER
Note: In Oracle Database XE Edition, RULE_CLASSIFIER and SVM_CLASSIFIER are not supported because the Data Mining option is not available. This is also true for KMEAN_CLUSTERING.

**RULE_CLASSIFIER**

Use the RULE_CLASSIFIER type for creating preferences for the query rule generating procedure, CTX_CLS.TRAIN and for CTXRULE creation. The rules generated with this type are essentially query strings and can be easily examined. The queries generated by this classifier can use the AND, NOT, or ABOUT operators. The WITHIN operator is supported for queries on field sections only.

This type has the following attributes:

*Table 2–44  RULE_CLASSIFIER Attributes*

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Data Type</th>
<th>Default</th>
<th>Min Value</th>
<th>Max Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>THRESHOLD</td>
<td>I</td>
<td>50</td>
<td>1</td>
<td>99</td>
<td>Specify threshold (in percentage) for rule generation. One rule is output only when its confidence level is larger than threshold.</td>
</tr>
<tr>
<td>MAX_TERMS</td>
<td>I</td>
<td>100</td>
<td>20</td>
<td>2000</td>
<td>For each class, a list of relevant terms is selected to form rules. Specify the maximum number of terms that can be selected for each class.</td>
</tr>
<tr>
<td>MEMORY_SIZE</td>
<td>I</td>
<td>500</td>
<td>10</td>
<td>4000</td>
<td>Specify memory usage for training in MB. Larger values improve performance.</td>
</tr>
<tr>
<td>NT_THRESHOLD</td>
<td>F</td>
<td>0.001</td>
<td>0</td>
<td>0.90</td>
<td>Specify a threshold for term selection. There are two thresholds guiding two steps in selecting relevant terms. This threshold controls the behavior of the first step. At this step, terms are selected as candidate terms for the further consideration in the second step. The term is chosen when the ratio of the occurrence frequency over the number of documents in the training set is larger than this threshold.</td>
</tr>
<tr>
<td>TERM_THRESHOLD</td>
<td>I</td>
<td>10</td>
<td>0</td>
<td>100</td>
<td>Specify a threshold as a percentage for term selection. This threshold controls the second step term selection. Each candidate term has a numerical quantity calculated to imply its correlation with a given class. The candidate term will be selected for this class only when the ratio of its quantity value over the maximum value for all candidate terms in the class is larger than this threshold.</td>
</tr>
</tbody>
</table>
Use the SVM_CLASSIFIER type for creating preferences for the rule generating procedure, CTX_CLS TRAIN, and for CTXRULE creation. This classifier type represents the Support Vector Machine method of classification and generates rules in binary format. Use this classifier type when you need high classification accuracy.

This type has the following attributes:

### Table 2–45 SVM_CLASSIFIER Attributes

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Data Type</th>
<th>Default</th>
<th>Min Value</th>
<th>Max Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAX_DOCTERMS</td>
<td>I</td>
<td>50</td>
<td>10</td>
<td>8192</td>
<td>Specify the maximum number of terms representing one document.</td>
</tr>
<tr>
<td>MAX_FEATURES</td>
<td>I</td>
<td>3,000</td>
<td>1</td>
<td>100,000</td>
<td>Specify the maximum number of distinct features.</td>
</tr>
<tr>
<td>THEME_ON</td>
<td>B</td>
<td>FALSE</td>
<td>NULL</td>
<td>NULL</td>
<td>Specify TRUE to use themes as features. Classification with themes requires an installed knowledge base. A knowledge base may or may not have been installed with Oracle Text. For more information on knowledge bases, see the Oracle Text Application Developer’s Guide.</td>
</tr>
<tr>
<td>TOKEN_ON</td>
<td>B</td>
<td>TRUE</td>
<td>NULL</td>
<td>NULL</td>
<td>Specify TRUE to use regular tokens as features.</td>
</tr>
<tr>
<td>STEM_ON</td>
<td>B</td>
<td>FALSE</td>
<td>NULL</td>
<td>NULL</td>
<td>Specify TRUE to use stemmed tokens as features. This only works when turning INDEX_STEM on for the lexer.</td>
</tr>
<tr>
<td>MEMORY_SIZE</td>
<td>I</td>
<td>500</td>
<td>10</td>
<td>4000</td>
<td>Specify approximate memory size in MB.</td>
</tr>
</tbody>
</table>
Cluster Types

This section describes the cluster types used for creating preferences for the CTX_CLS.CLUSTERING procedure.

Note: In Oracle Database XE Edition, KMEAN_CLUSTERING is not supported because the Data Mining option is not available. This is also true for RULE_CLASSIFIER and SVM_CLASSIFIER.

See Also: For more information about clustering, see "CLUSTERING" in Chapter 6, "CTX_CLS Package" as well as the Oracle Text Application Developer’s Guide

KMEAN_CLUSTERING

This clustering type has the following attributes:

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Data Type</th>
<th>Default</th>
<th>Min Value</th>
<th>Max Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAX_DOCTERMS</td>
<td>I</td>
<td>50</td>
<td>10</td>
<td>8192</td>
<td>Specify the maximum number of distinct terms representing one document.</td>
</tr>
<tr>
<td>MAX_FEATURES</td>
<td>I</td>
<td>3,000</td>
<td>1</td>
<td>500,000</td>
<td>Specify the maximum number of distinct features.</td>
</tr>
<tr>
<td>THEME_ON</td>
<td>B</td>
<td>FALSE</td>
<td>NULL</td>
<td>NULL</td>
<td>Specify TRUE to use themes as features. Clustering with themes requires an installed knowledge base. A knowledge base may or may not have been installed with Oracle Text. For more information on knowledge bases, see Oracle Text Application Developer’s Guide.</td>
</tr>
</tbody>
</table>
Stoplists

Stoplists identify the words in your language that are not to be indexed. In English, you can also identify stopthemes that are not to be indexed. By default, the system indexes text using the system-supplied stoplist that corresponds to your database language.

Oracle Text provides default stoplists for most common languages including English, French, German, Spanish, Chinese, Dutch, and Danish. These default stoplists contain only stopwords.

**See Also:** For more information about the supplied default stoplists, see Appendix E, "Oracle Text Supplied Stoplists"

Multi-Language Stoplists

You can create multi-language stoplists to hold language-specific stopwords. A multi-language stoplist is useful when you use the MULTI_LEXER to index a table that contains documents in different languages, such as English, German.

To create a multi-language stoplist, use the CTX_DLL.CREATE_STOPLIST procedure and specify a stoplist type of MULTI_STOPLIST. Add language specific stopwords with CTX_DDL.ADD_STOPWORD.

At indexing time, the language column of each document is examined, and only the stopwords for that language are eliminated. At query time, the session language setting determines the active stopwords, like it determines the active lexer when using the multi-lexer.
Creating Stoplists

Create your own stoplists using the CTX_DLL.CREATE_STOPLIST procedure. With this procedure you can create a BASIC_STOPLIST for single language stoplist, or you can create a MULTI_STOPLIST for a multi-language stoplist.

When you create your own stoplist, you must specify it in the parameter clause of CREATE INDEX.

To create stoplists for Chinese or Japanese languages, use the CHINESE_LEXER or JAPANESE_LEXER respectively, and update the appropriate lexicon to be @contained_such_stopwords.

Modifying the Default Stoplist

The default stoplist is always named CTXSYS.DEFAULT_STOPLIST. Use the following procedures to modify this stoplist:

- CTX_DDL.ADD_STOPWORD
- CTX_DDL.REMOVE_STOPWORD
- CTX_DDL.ADD_STOPTHEME
- CTX_DDL.ADD_STOPCLASS

When you modify CTXSYS.DEFAULT_STOPLIST with the CTX_DDL package, you must re-create your index for the changes to take effect.

Dynamic Addition of Stopwords

You can add stopwords dynamically to a default or custom stoplist with ALTER INDEX. When you add a stopword dynamically, you need not re-index, because the word immediately becomes a stopword and is removed from the index.

---

Note: Even though you can dynamically add stopwords to an index, you cannot dynamically remove stopwords. To remove a stopword, you must use CTX_DDL.REMOVE_STOPWORD, drop your index and re-create it.

---

See Also: "ALTER INDEX" in Chapter 1, "Oracle Text SQL Statements and Operators"

System-Defined Preferences

When you install Oracle Text, some indexing preferences are created. You can use these preferences in the parameter clause of CREATE INDEX or define your own.

The default index parameters are mapped to some of the system-defined preferences described in this section.

See Also: For more information about default index parameters, see "Default Index Parameters" on page 2-88

System-defined preferences are divided into the following categories:

- Data Storage
- Filter
Data Storage

This section discusses the types associated with data storage preferences.

**CTXSYS.DEFAULT_DATASTORE**
This preference uses the DIRECT_DATASTORE type. Use this preference to create indexes for text columns in which the text is stored directly in the column.

**CTXSYS.FILE_DATASTORE**
This preference uses the FILE_DATASTORE type.

**CTXSYS.URL_DATASTORE**
This preference uses the URL_DATASTORE type.

Filter

This section discusses the types associated with filtering preferences.

**CTXSYS.NULL_FILTER**
This preference uses the NULL_FILTER type.

**CTXSYS.AUTO_FILTER**
This preference uses the AUTO_FILTER type.

Lexer

This section discusses the types associated with lexer preferences.

**CTXSYS.DEFAULT_LEXER**
The default lexer depends on the language used at install time. The following sections describe the default settings for CTXSYS.DEFAULT_LEXER for each language.

**American and English Language Settings** If your language is English, this preference uses the BASIC_LEXER with the index_themes attribute disabled.

**Danish Language Settings** If your language is Danish, this preference uses the BASIC_LEXER with the following option enabled:
- Alternate spelling (alternate_spelling attribute set to DANISH)

**Dutch Language Settings** If your language is Dutch, this preference uses the BASIC_LEXER with the following options enabled:
- composite indexing (composite attribute set to DUTCH)
German and German DIN Language Settings  If your language is German, then this preference uses the BASIC_LEXER with the following options enabled:
- Case-sensitive indexing (mixed_case attribute enabled)
- Composite indexing (composite attribute set to GERMAN)
- Alternate spelling (alternate_spelling attribute set to GERMAN)

Finnish, Norwegian, and Swedish Language Settings  If your language is Finnish, Norwegian, or Swedish, this preference uses the BASIC_LEXER with the following option enabled:
- Alternate spelling (alternate_spelling attribute set to SWEDISH)

Japanese Language Settings  If you language is Japanese, this preference uses the JAPANESE_VGRAM_LEXER.

Korean Language Settings  If your language is Korean, this preference uses the KOREAN_MORPH_LEXER. All attributes for the KOREAN_MORPH_LEXER are enabled.

Chinese Language Settings  If your language is Simplified or Traditional Chinese, this preference uses the CHINESE_VGRAM_LEXER.

Other Languages  For all other languages not listed in this section, this preference uses the BASIC_LEXER with no attributes set.

See Also:  To learn more about these options, see "BASIC_LEXER" on page 2-39

CTXSYS.DEFAULT_EXTRACT_LEXER
This preference uses AUTO_LEXER with the following options:
- alternate_spelling is NONE
- base_letter is NO
- mixed_case is YES

CTXSYS.BASIC_LEXER
This preference uses the BASIC_LEXER.

Section Group
This section discusses the types associated with section group preferences.

CTXSYS.NULL_SECTION_GROUP
This preference uses the NULL_SECTION_GROUP type.

CTXSYS.HTML_SECTION_GROUP
This preference uses the HTML_SECTION_GROUP type.

CTXSYS.AUTO_SECTION_GROUP
This preference uses the AUTO_SECTION_GROUP type.
**CTXSYS.PATH_SECTION_GROUP**
This preference uses the PATH_SECTION_GROUP type.

**Stoplist**
This section discusses the types associated with stoplist preferences.

**CTXSYS.DEFAULT_STOPLIST**
This stoplist preference defaults to the stoplist of your database language.

*See Also:* For a complete list of the stop words in the supplied stoplists, see Appendix E, "Oracle Text Supplied Stoplists"

**CTXSYS.EMPTY_STOPLIST**
This stoplist has no words.

**Storage**
This section discusses the types associated with storage preferences.

**CTXSYS.DEFAULT_STORAGE**
This storage preference uses the BASIC_STORAGE type.

**Wordlist**
This section discusses the types associated with wordlist preferences.

**CTXSYS.DEFAULT_WORDLIST**
This preference uses the language stemmer for your database language. If your language is not listed in Table 2–40 on page 2-68, then this preference defaults to the NULL stemmer and the GENERIC fuzzy matching attribute.

**System Parameters**
This section describes the Oracle Text system parameters, which are divided into the following categories:

- General System Parameters
- Default Index Parameters

**General System Parameters**
When you install Oracle Text, in addition to the system-defined preferences, the following system parameters are set:

*Table 2–47 General System Parameters*

<table>
<thead>
<tr>
<th>System Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAX_INDEX_MEMORY</td>
<td>This is the maximum indexing memory that can be specified in the parameter clause of CREATE INDEX and ALTER INDEX. The maximum value for this parameter is 2 GB -1.</td>
</tr>
<tr>
<td>DEFAULT_INDEX_MEMORY</td>
<td>This is the default indexing memory used with CREATE INDEX and ALTER INDEX.</td>
</tr>
</tbody>
</table>
View system defaults by querying the CTX_PARAMETERS view. Change defaults using the CTX_ADM.SET_PARAMETER procedure.

Default Index Parameters

This section describes the index parameters that you can use when you create CONTEXT and CTXCAT indexes.

CONTEXT Index Parameters

The following default parameters are used when you create a CONTEXT index and do not specify preferences in the parameter clause of CREATE INDEX. Each default parameter names a system-defined preference to use for data storage, filtering, lexing, and so on.

Table 2–48 Default CONTEXT Index Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Used When</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFAULT_DATASTORE</td>
<td>No datastore preference specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_DATASTORE</td>
</tr>
<tr>
<td>DEFAULT_FILTER_FILE</td>
<td>No filter preference specified in parameter clause of CREATE INDEX, and either of the following conditions is true:</td>
<td>CTXSYS.AUTO_FILTER</td>
</tr>
<tr>
<td></td>
<td>■ Your files are stored in external files (BFILES) or</td>
<td></td>
</tr>
<tr>
<td></td>
<td>■ Specify a datastore preference that uses FILE_DATASTORE</td>
<td></td>
</tr>
<tr>
<td>DEFAULT_FILTER_BINARY</td>
<td>No filter preference specified in parameter clause of CREATE INDEX, and Oracle Text detects that the text column datatype is RAW, LONG RAW, or BLOB.</td>
<td>CTXSYS.AUTO_FILTER</td>
</tr>
<tr>
<td>DEFAULT_FILTER_TEXT</td>
<td>No filter preference specified in parameter clause of CREATE INDEX, and Oracle Text detects that the text column datatype is either LONG, VARCHAR2, VARCHAR, CHAR, or CLOB.</td>
<td>CTXSYS.NULL_FILTER</td>
</tr>
</tbody>
</table>
Table 2–48 (Cont.) Default CONTEXT Index Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Used When</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFAULT_SECTION_HTML</td>
<td>No section group specified in parameter clause of CREATE INDEX, and when either of the following conditions is true:</td>
<td>CTXSYS.HTML_SECTION_GROUP</td>
</tr>
<tr>
<td></td>
<td>■ Your datastore preference uses URL_DATASTORE or</td>
<td></td>
</tr>
<tr>
<td></td>
<td>■ Your filter preference uses AUTO_FILTER.</td>
<td></td>
</tr>
<tr>
<td>DEFAULT_SECTION_TEXT</td>
<td>No section group specified in parameter clause of CREATE INDEX, and when you do not use either URL_DATASTORE or AUTO_FILTER.</td>
<td>CTXSYS.NULL_SECTION_GROUP</td>
</tr>
<tr>
<td>DEFAULT_STORAGE</td>
<td>No storage preference specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_STORAGE</td>
</tr>
<tr>
<td>DEFAULT_LEXER</td>
<td>No lexer preference specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_LEXER</td>
</tr>
<tr>
<td>DEFAULT_STOPLIST</td>
<td>No stoplist specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_STOPLIST</td>
</tr>
<tr>
<td>DEFAULT_WORDLIST</td>
<td>No wordlist preference specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_WORDLIST</td>
</tr>
</tbody>
</table>

Table 2–49 Default CTXCAT Index Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Used When</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFAULT_CTXCAT_INDEX_SET</td>
<td>No index set specified in parameter clause of CREATE INDEX.</td>
<td>n/a</td>
</tr>
<tr>
<td>DEFAULT_CTXCAT_STORAGE</td>
<td>No storage preference specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_STORAGE</td>
</tr>
<tr>
<td>DEFAULT_CTXCAT_LEXER</td>
<td>No lexer preference specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_LEXER</td>
</tr>
<tr>
<td>DEFAULT_CTXCAT_STOPLIST</td>
<td>No stoplist specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_STOPLIST</td>
</tr>
</tbody>
</table>

CTXCAT Index Parameters

The following default parameters are used when you create a CTXCAT index with CREATE INDEX and do not specify any parameters in the parameter string. The CTXCAT index supports only the index set, lexer, storage, stoplist, and wordlist parameters. Each default parameter names a system-defined preference.
System Parameters

Table 2–49 (Cont.) Default CTXCAT Index Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Used When</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFAULT_CTXCAT_WORDLIST</td>
<td>No wordlist preference specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_CTXCAT_WORDLIST</td>
</tr>
<tr>
<td></td>
<td>Note that while you can specify a wordlist preference for CTXCAT indexes, most of the attributes do not apply, because the catsearch query language does not support wildcarding, fuzzy, and stemming. The only attribute that is useful is PREFIX_INDEX for Japanese data.</td>
<td></td>
</tr>
</tbody>
</table>

CTXRULE Index Parameters

Table 2–50 lists the default parameters that are used when you create a CTXRULE index with CREATE INDEX and do not specify any parameters in the parameter string. The CTXRULE index supports only the lexer, storage, stoplist, and wordlist parameters. Each default parameter names a system-defined preference.

Table 2–50 Default CTXRULE Index Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Used When</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFAULT_CTXRULE_LEXER</td>
<td>No lexer preference specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_CTXRULE_LEXER</td>
</tr>
<tr>
<td>DEFAULT_CTXRULE_STORAGE</td>
<td>No storage preference specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_CTXRULE_STORAGE</td>
</tr>
<tr>
<td>DEFAULT_CTXRULE_STOPLIST</td>
<td>No stoplist specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_CTXRULE_STOPLIST</td>
</tr>
<tr>
<td>DEFAULT_CTXRULE_WORDLIST</td>
<td>No wordlist preference specified in parameter clause of CREATE INDEX.</td>
<td>CTXSYS.DEFAULT_CTXRULE_WORDLIST</td>
</tr>
<tr>
<td>DEFAULT_CLASSIFIER</td>
<td>No classifier preference is specified in parameter clause.</td>
<td>RULE_CLASSIFIER</td>
</tr>
</tbody>
</table>

CTXRULE Index Limitations

The CTXRULE index does not support the following query operators:

- Fuzzy
- Soundex

It also does not support the following BASIC_WORDLIST attributes:

- SUBSTRING_INDEX
- PREFIX_INDEX

Viewing Default Values

View system defaults by querying the CTX_PARAMETERS view. For example, to see all parameters and values, enter the following statement:
SQL> SELECT par_name, par_value from ctx_parameters;

**Changing Default Values**
Change a default value using the CTX_ADM.SET_PARAMETER procedure to name another custom or system-defined preference to use as default.
This chapter describes operator precedence and provides descriptions, syntax, and examples for every CONTAINS operator. The following topics are covered:

- Operator Precedence
- ABOUT
- ACCUMulate (, )
- AND (&)
- Broader Term (BT, BTG, BTP, BTI)
- DEFINEMERGE
- DEFINESCORE
- EQUIValence (=)
- Fuzzy
- HASPATH
- INPATH
- MDATA
- MINUS (-)
- MNOT
- Narrower Term (NT, NTG, NTP, NTI)
- NDATA
- NEAR (;)
- NOT (~)
- OR (|)
- Preferred Term (PT)
- Related Term (RT)
- SDATA
- soundex (!)
- stem ($)
- Stored Query Expression (SQE)
- SYNonym (SYN)
Operator Precedence

Operator precedence determines the order in which the components of a query expression are evaluated. Text query operators can be divided into two sets of operators that have their own order of evaluation. These two groups are described later as Group 1 and Group 2.

In all cases, query expressions are evaluated in order from left to right according to the precedence of their operators. Operators with higher precedence are applied first. Operators of equal precedence are applied in order of their appearance in the expression from left to right.

Group 1 Operators

Within query expressions, the Group 1 operators have the following order of evaluation from highest precedence to lowest:

1. EQUIValence (=)
2. NEAR (;)
3. weight (*), threshold (>)
4. MINUS (-)
5. NOT (~)
6. MNOT
7. WITHIN
8. AND (&)
9. OR (|)
10. ACCUMulate (, )

Group 2 Operators and Characters

Within query expressions, the Group 2 operators have the following order of evaluation from highest to lowest:

1. Wildcard Characters
2. stem ($)
3. Fuzzy
4. soundex (!)
Procedural Operators

Other operators not listed under Group 1 or Group 2 are procedural. These operators have no sense of precedence attached to them. They include the SQE and thesaurus operators.

Precedence Examples

<table>
<thead>
<tr>
<th>Query Expression</th>
<th>Order of Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1</td>
<td>w2 &amp; w3</td>
</tr>
<tr>
<td>w1 &amp; w2</td>
<td>w3</td>
</tr>
<tr>
<td>?w1, w2</td>
<td>w3 &amp; w4</td>
</tr>
<tr>
<td>abc = def ghi &amp; jkl = mno</td>
<td>((abc = def ghi) &amp; (jkl=mno))</td>
</tr>
<tr>
<td>dog and cat WITHIN body</td>
<td>dog and (cat WITHIN body)</td>
</tr>
</tbody>
</table>

In the first example, because AND has a higher precedence than OR, the query returns all documents that contain w1 and all documents that contain both w2 and w3.

In the second example, the query returns all documents that contain both w1 and w2 and all documents that contain w3.

In the third example, the fuzzy operator is first applied to w1, then the AND operator is applied to arguments w3 and w4, then the OR operator is applied to term w2 and the results of the AND operation, and finally, the score from the fuzzy operation on w1 is added to the score from the OR operation.

The fourth example shows that the equivalence operator has higher precedence than the AND operator.

The fifth example shows that the AND operator has lower precedence than the WITHIN operator.

Altering Precedence

Precedence is altered by grouping characters as follows:

- Within parentheses, expansion or execution of operations is resolved before other expansions regardless of operator precedence.
- Within parentheses, precedence of operators is maintained during evaluation of expressions.
- Within parentheses, expansion operators are not applied to expressions unless the operators are also within the parentheses.

See Also: "Grouping Characters" in Chapter 4, "Special Characters in Oracle Text Queries"
ABOUT

General Behavior

Use the ABOUT operator to return documents that are related to a query term or phrase. In English and French, ABOUT enables you to query on concepts, even if a concept is not actually part of a query. For example, an ABOUT query on heat might return documents related to temperature, even though the term temperature is not part of the query.

In other languages, using ABOUT will often increase the number of returned documents and may improve the sorting order of results. For all languages, Oracle Text scores results for an ABOUT query with the most relevant document receiving the highest score.

English and French Behavior

In English and French, use the ABOUT operator to query on concepts. The system looks up concept information in the theme component of the index. Create a theme component to your index by setting the INDEX_THEMES BASIC_LEXER attribute to YES.

Note: You need not have a theme component in the index to enter ABOUT queries in English and French. However, having a theme component in the index yields the best results for ABOUT queries.

Oracle Text retrieves documents that contain concepts that are related to your query word or phrase. For example, if you enter an ABOUT query on California, the system might return documents that contain the terms Los Angeles and San Francisco, which are cities in California. The document need not contain the term California to be returned in this ABOUT query.

The word or phrase specified in your ABOUT query need not exactly match the themes stored in the index. Oracle Text normalizes the word or phrase before performing lookup in the index.

You can use the ABOUT operator with the CONTAINS and CATSEARCH SQL operators. In the case of CATSEARCH, you must use query templating with the CONTEXT grammar to query on the indexed themes. See ABOUT Query with CATSEARCH in the Examples section.
Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>about(phrase)</td>
<td>In all languages, increases the number of relevant documents returned for the same query without the ABOUT operator. The phrase parameter can be a single word or a phrase, or a string of words in free text format.</td>
</tr>
<tr>
<td></td>
<td>In English and French, returns documents that contain concepts related to phrase, provided the BASIC_LEXER INDEX_THEMES attribute is set to YES at index time.</td>
</tr>
<tr>
<td></td>
<td>The score returned is a relevance score.</td>
</tr>
<tr>
<td></td>
<td>Oracle Text ignores any query operators that are included in phrase.</td>
</tr>
<tr>
<td></td>
<td>If your index contains only theme information, an ABOUT operator and operand must be included in your query on the text column or else Oracle Text returns an error.</td>
</tr>
<tr>
<td></td>
<td>The phrase you specify cannot be more than 4000 characters.</td>
</tr>
</tbody>
</table>

Case-Sensitivity

ABOUT queries give the best results when your query is formulated with proper case. This is because the normalization of your query is based on the knowledge catalog which is case-sensitive.

However, you need not type your query in exact case to obtain results from an ABOUT query. The system does its best to interpret your query. For example, if you enter a query of CISCO and the system does not find this in the knowledge catalog, the system might use Cisco as a related concept for look-up.

Improving ABOUT Results

The ABOUT operator uses the supplied knowledge base in English and French to interpret the phrase you enter. Your ABOUT query therefore is limited to knowing and interpreting the concepts in the knowledge base.

Improve the results of your ABOUT queries by adding your application-specific terminology to the knowledge base.

See Also: "Extending the Knowledge Base" in Chapter 15, "Oracle Text Utilities"

Limitations

The phrase you specify in an ABOUT query cannot be more than 4000 characters.

Examples

Single Words

To search for documents that are about soccer, use the following syntax:

‘about{soccer}'

Phrases

Further refine the query to include documents about soccer rules in international competition by entering the phrase as the query term:

‘about{soccer rules in international competition}’
In this English example, Oracle Text returns all documents that have themes of *soccer*, *rules*, or *international competition*.

In terms of scoring, documents which have all three themes will generally score higher than documents that have only one or two of the themes.

### Unstructured Phrases
You can also query on unstructured phrases, such as the following:

'about(japanese banking investments in indonesia)'

### Combined Queries
Use other operators, such as **AND** or **NOT**, to combine ABOUT queries with word queries. For example, enter the following combined ABOUT and word query:

'about(dogs) and cat'

Combine an ABOUT query with another ABOUT query as follows:

'about(dogs) not about(labradors)'

---

**Note:** You cannot combine ABOUT with the **WITHIN** operator, as for example 'ABOUT (xyz) WITHIN abc'.

---

### ABOUT Query with CATSEARCH
Enter ABOUT queries with CATSEARCH using the query template method with grammar set to **CONTEXT** as follows:

```sql
select pk||' ==> '||text from test
where catsearch(text,
    '<query>
        <textquery grammar="context">
            about(California)
        </textquery>
        <score datatype="integer"/>
    </query>',"')>0
order by pk;
```
**ACCUMulate (, )**

Use the ACCUM operator to search for documents that contain at least one occurrence of any query terms, with the returned documents ranked by a cumulative score based on how many query terms are found (and how frequently).

**Syntax**

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>term1,term2</code></td>
<td>Returns documents that contain <code>term1</code> or <code>term2</code>. Ranks documents according to document term weight, with the highest scores assigned to documents that have the highest total term weight.</td>
</tr>
<tr>
<td><code>term1 ACCUM term2</code></td>
<td></td>
</tr>
</tbody>
</table>

**ACCUMulate Scoring**

ACCUMulate first scores documents on how many query terms a document matches. A document that matches more terms will always score higher than a document that matches fewer terms, even if the terms appear more frequently in the latter. In other words, if you search for `dog ACCUM cat`, you'll find that

the dog played with the cat

scores higher than

the big dog played with the little dog while a third dog ate the dog food

Scores are divided into ranges. In a two-term ACCUM, hits that match both terms will always score between 51 and 100, whereas hits matching only one of the terms will score between 1 and 50. Likewise, for a three-term ACCUM, a hit matching one term will score between 1 and 33; a hit matching two terms will score between 34 and 66, and a hit matching all three terms will score between 67 and 100. *Within these ranges*, normal scoring algorithms apply.

**See Also:** Appendix F, "The Oracle Text Scoring Algorithm" for more information on how scores are calculated

You can assign different weights to different terms. For example, in a query of the form

soccer, Brazil*3

the term Brazil is weighted three times as heavily as soccer. Therefore, the document

people play soccer because soccer is challenging and fun

will score lower than

Brazil is the largest nation in South America

but both documents will rank below

soccer is the national sport of Brazil

Note that a query of `soccer ACCUM Brazil*3` is equivalent to `soccer ACCUM Brazil ACCUM Brazil ACCUM Brazil`. Because each query term Brazil is considered independent, the entire query is scored as though it has four terms, not two, and thus has four scoring ranges. The first Brazil-and-soccer example document shown above scores in the first range (1-25), the second scores in the third range (51-75), and the
third scores in the fourth range (76-100). (No document scores in the second range, because any document with Brazil in it will be considered to match at least three query terms.)

Example

set serveroutput on;
DROP TABLE accumtbl;
CREATE TABLE accumtbl (id NUMBER, text VARCHAR2(4000) );

INSERT INTO accumtbl VALUES ( 1, 'the little dog played with the big dog while the other dog ate the dog food');
INSERT INTO accumtbl values (2, 'the cat played with the dog');

CREATE INDEX accumtbl_idx ON accumtbl (text) indextype is ctxsys.context;

PROMPT dog ACCUM cat
SELECT SCORE(10) FROM accumtbl WHERE CONTAINS (text, 'dog ACCUM cat', 10) > 0;

PROMPT dog*3 ACCUM cat
SELECT SCORE(10) FROM accumtbl WHERE CONTAINS (text, 'dog*3 ACCUM cat', 10) > 0;

This produces the following output. Note that the document with both dog and cat scores highest.

dog ACCUM cat
<table>
<thead>
<tr>
<th>ID</th>
<th>SCORE(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>52</td>
</tr>
</tbody>
</table>

dog*3 ACCUM cat
<table>
<thead>
<tr>
<th>ID</th>
<th>SCORE(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>53</td>
</tr>
<tr>
<td>2</td>
<td>76</td>
</tr>
</tbody>
</table>

Related Topics

See also weight (*) on page 3-56
Use the AND operator to search for documents that contain at least one occurrence of each of the query terms.

**Syntax**

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>term1&amp;term2</code></td>
<td>Returns documents that contain <code>term1</code> and <code>term2</code>. Returns the minimum score of its operands. All query terms must occur; lower score taken.</td>
</tr>
<tr>
<td><code>term1 and term2</code></td>
<td></td>
</tr>
</tbody>
</table>

**Example**

To obtain all the documents that contain the terms *blue* and *black* and *red*, enter the following query:

`‘blue & black & red’`

In an AND query, the score returned is the score of the lowest query term. In this example, if the three individual scores for the terms *blue*, *black*, and *red* is 10, 20 and 30 within a document, the document scores 10.

**Related Topics**

**See Also:** The AND operator returns documents that contain *all* of the query terms, while OR operator returns documents that contain *any* of the query terms. See "OR (|)" on page 3-41.
Broader Term (BT, BTG, BTP, BTI)

Use the broader term operators (BT, BTG, BTP, BTI) to expand a query to include the term that has been defined in a thesaurus as the broader or higher level term for a specified term. They can also expand the query to include the broader term for the broader term and the broader term for that broader term, and so on up through the thesaurus hierarchy.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT(term[(qualifier)][,level][,thes])</td>
<td>Expands a query to include the term defined in the thesaurus as a broader term for term.</td>
</tr>
<tr>
<td>BTG(term[(qualifier)][,level][,thes])</td>
<td>Expands a query to include all terms defined in the thesaurus as broader generic terms for term.</td>
</tr>
<tr>
<td>BTP(term[(qualifier)][,level][,thes])</td>
<td>Expands a query to include all the terms defined in the thesaurus as broader partitive terms for term.</td>
</tr>
<tr>
<td>BTI(term[(qualifier)][,level][,thes])</td>
<td>Expands a query to include all the terms defined in the thesaurus as broader instance terms for term.</td>
</tr>
</tbody>
</table>

**term**

Specify the operand for the broader term operator. Oracle Text expands term to include the broader term entries defined for the term in the thesaurus specified by thes. For example, if you specify BTG(dog), the expansion includes only those terms that are defined as broader term generic for dog. You cannot specify expansion operators in the term argument.

The number of broader terms included in the expansion is determined by the value for level.

**qualifier**

Specify a qualifier for term, if term is a homograph (word or phrase with multiple meanings, but the same spelling) that appears in two or more nodes in the same hierarchy branch of thes.

If a qualifier is not specified for a homograph in a broader term query, the query expands to include the broader terms of all the homographic terms.

**level**

Specify the number of levels traversed in the thesaurus hierarchy to return the broader terms for the specified term. For example, a level of 1 in a BT query returns the broader term entry, if one exists, for the specified term. A level of 2 returns the broader term entry for the specified term, as well as the broader term entry, if one exists, for the broader term.

The level argument is optional and has a default value of one (1). Zero or negative values for the level argument return only the original query term.

**thes**

Specify the name of the thesaurus used to return the expansions for the specified term. The thes argument is optional and has a default value of DEFAULT. A thesaurus named DEFAULT must exist in the thesaurus tables if you use this default value.
Examples

The following query returns all documents that contain the term tutorial or the BT term defined for tutorial in the DEFAULT thesaurus:

'BT(tutorial)'

When you specify a thesaurus name, you must also specify level as in:

'BT(tutorial, 2, mythes)'

Broader Term Operator on Homographs

If machine is a broader term for crane (building equipment) and bird is a broader term for crane (waterfowl) and no qualifier is specified for a broader term query, the query

BT(crane)

expands to:

' (crane) or (machine) or (bird)'

If waterfowl is specified as a qualifier for crane in a broader term query, the query

BT(crane{(waterfowl)})

expands to the query:

' (crane) or (bird)'

Note: When specifying a qualifier in a broader or narrower term query, the qualifier and its notation (parentheses) must be escaped, as is shown in this example.

Related Topics

Browse a thesaurus using procedures in the CTX_THES package.

See Also:  CTX_THES.BT in Chapter 13, "CTX_THES Package" for more information on browsing the broader terms in your thesaurus.
DEFINEMERGE

Use the DEFINEMERGE operator to define how the score of child nodes of the AND and OR should be merged. The DEFINEMERGE operator can be used as operand(s) of any operators that allow AND or OR as operands. The score can be merged in three ways: picking the minimum value, picking the maximum value, or calculating the average score of all child nodes.

Use DEFINESCORE before using DEFINEMERGE.

Syntax

DEFINEMERGE ( ( (text_query1), (text_query2), ...) , operator, merge_method )

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>text_query1,2 ...</td>
<td>Defines the search criteria. These parameters can have any value that is valid for the AND/OR operator.</td>
</tr>
<tr>
<td>operator</td>
<td>Defines the relationship between the two text_query parameters.</td>
</tr>
<tr>
<td>merge_method</td>
<td>Defines how the score of the text_query should be merged. Possible values: MIN, MAX, AVG, ADD</td>
</tr>
</tbody>
</table>

Example

Example 3–1 DEFINEMERGE and text_query

The following examples show only the text_query part of a CONTAINS query:

'DEFINEMERGE ( ((dog), (cat)), OR, AVG)'

Queries for the term "dog" or "cat," using the average relevance score of both terms as the merged score.

'DEFINEMERGE (((dog , cat) , (blue or black)), AND, MIN)'

Queries for the expression "dog ACCUM cat" and "blue OR black," using the default scoring schemes and then using the minimum score of the two as the merged-score.

'DEFINEMERGE( ((DEFINESCORE(dog, DISCRETE)) , (cat)), AND, MAX)'

Queries for the term "dog" using the DISCRETE scoring, and for the term "cat" using the default relevant scoring, and then using the maximum score of the two as the merged-score.

Related Topic

DEFINESCORE on page 3-13.
Use the `DEFINESCORE` operator to define how a term or phrase, or a set of term equivalences will be scored. The definition of a scoring expression can consist of an arithmetic expression of predefined scoring components and numeric literals.

`DEFINEMERGE` can be used after `DEFINESCORE`.

**Syntax**

```sql
DEFINESCORE (query_term, scoring_expression)
```

**query_term**
The query term or phrase. Expressions containing the following operators are also allowed:

- ABOUT EQUIV(=)
- Fuzzy Soundex (!)
- Stem ($) Wildcards (% _)
- SDATA MDATA

**scoring_expression**
An arithmetic expression that describes how the `query_term` should be scored. This operand is a string that contains the following components:

- Arithmetic operators: + - * /. The precedence is multiplication and division (*, /) first before addition and subtraction (+, -).
- Grouping operators: (). Parentheses can be used to alter the precedence of the arithmetic operators.
- Absolute function: `ABS(n)` returns the absolute value of `n`; where `n` is any expression that returns a number.
- Logarithmic function: `LOG(n)` returns the base-10 logarithmic value of `n`; where `n` is any expression that returns a number.
- Predefined scoring components: Each of the following scoring components returns a value of 0 - 100, depending on different criteria:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISCRETE</td>
<td>If the term exists in the document, score = 100. Otherwise, score = 0.</td>
</tr>
<tr>
<td>OCCURRENCE</td>
<td>Score based on the number of occurrences.</td>
</tr>
<tr>
<td>RELEVANCE</td>
<td>Score based on the document’s relevance.</td>
</tr>
<tr>
<td>COMPLETION</td>
<td>Score based on coverage. Documents will score higher if the ratio between the number of matching terms and the number of all terms in the section (counting stop words) is higher. The COMPLETION scoring is only applicable when used with the WITHIN operator to search in zone sections.</td>
</tr>
</tbody>
</table>
DEFINESCORE

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGNORE</td>
<td>Ignore the scoring of this term. This component should be used alone. Otherwise, the query will return a syntax error. If the scoring of the only term in the query is set to IGNORE, then all the matching documents should be returned with the same score of 100.</td>
</tr>
</tbody>
</table>

**Note:** For numeric literals, any number literal can be used that conforms to the SQL pattern of number literal, and is within the range of the double precision floating point (\(-3.4e38 \) to \(3.4e38\)).

### scoring_expression Syntax

\[
\begin{align*}
\text{<Exp>} & : = & \text{<Exp>} + \text{<Term>} & | & \text{<Exp>} - \text{<Term>} & | & \text{<Term>} \\
\text{<Term>} & : = & \text{<Term>} * \text{<Factor>} & | & \text{<Term>} / \text{<Factor>} & | & \text{<Factor>} \\
\text{<Factor>} & : = & \langle\text{NumericLiterals} \rangle & | & \text{DISCRETE} & | & \text{OCCURRENCE} & | & \text{RELEVANCE} & | & \text{COMPLETION} & | & \text{IGNORE} & | & (\text{<Exp>}) & | & -\text{<Factor>} & | & \text{Abs}(\text{<Exp>}) & | & \text{Log}(\text{<Exp>})
\end{align*}
\]

### Examples

- `'DEFINESCORE (dog, OCCURRENCE)’`

  Queries for the word `dog`, and scores each document using the occurrence score. Returns the score as integer.

- `'DEFINESCORE (Labradors are big dog, RELEVANCE)’`

  Queries for the phrase `Labradors are big dogs`, and scores each document using the relevance score.

- `'cat and DEFINESCORE (dog, IGNORE)’`

  Queries for the words `dog` and `cat`, using only the default relevance score of `cat` as the overall score of the document. Returns the score as integer.

- `'DEFINESCORE (dog, IGNORE)’`

  Queries for the word `dog`, and returns all documents with the word `dog`. The result is the same as if all documents get a score of 100. Returns the score as integer.

- `'DEFINESCORE (dog, ABS (100-RELEVANCE))’`

  Queries for the word `dog`, and scores each document using the absolute value of 100 minus the relevance score. Returns the score as integer.

- `'cat and DEFINESCORE (dog, RELEVANCE*5 - OCCURRENCE)’`

  Returns a syntax error: Two pre-defined components are used.

When DEFINESCORE is used with query templates, the scoring_expression overrides the values specified by the template. The following example queries for "dog" and "cat," scores "cat" using OCCURRENCE(COUNT) and scores "dog" based on RELEVANCE.

```xml
<textquery grammar="CONTEXT" lang="english">
  DEFINESCORE(dog, RELEVANCE) and cat
</textquery>

<score datatype="INTEGER" algorithm="COUNT"/>
</query>
```
**Limitations**

- If the `ABOUT` operator is used in `query_term`, the `OCCURRENCE` and `COMPLETION` scoring will not be applicable. If used, the query will return a syntax error.

- The `IGNORE` score cannot be used as right hand of the minus operator. If used, then a syntax error will occur.

- The `COMPLETION` score is only applicable if the `DEFINESCORE` is used with a `WITHIN` operator to search in zone sections, for example:
  
  `DEFINESCORE (dog, COMPLETION) within zonesection`
  
  otherwise, the query will return a syntax error.

- For the left hand operand of `WITHIN`:
  - All nodes must use the same predefined-scoring component. (If not specified, then the predefined scoring is `RELEVANCE`.)
  - If the nodes use `DISCRETE` or `COMPLETION`, then only the `AND` and `OR` operator is allowed as the left hand children of `WITHIN`.
  - If the nodes use `DISCRETE` or `COMPLETION`, then `WITHIN` will use the max score of all section instances as the score.
  - If the nodes use `RELEVANCE` or `OCCURRENCE`, then `WITHIN` will use the summation of the score of all section instances as the score.

- Only one predefined scoring component can be used in the `scoring_expression` at one time. If more than one predefined scoring component is used, then a syntax error will occur.

**See Also:** *Oracle Database SQL Language Reference*

**Notes**

- The `DEFINESCORE` operator, the absolute function, the logarithmic function, and the pre-defined scoring components are case-insensitive.

- The `query_term` and the `scoring_expression` parameters are mandatory.

- The final score of the `DEFINESCORE` operator will be truncated to be in the 0 – 100 range. If the data type is `INTEGER`, then the score is rounded up.

- The intermediate data type of the scoring value is a double precision float. As a result, the value is limited to be in the $-3.4\times10^{38}$ to $3.4\times10^{38}$ range. If the intermediate scoring of any document exceeds the value, then the score will be truncated. If an integer scoring is required, then the score will always be rounded up after the score is calculated.

- The `DEFINESCORE` operator can be used as an operand of the following operators:
  - `AND`
  - `NOT`
  - `INPATH`
  - `THRESHOLD`
  - `WITHIN`
  - `SQE`
  - `OR`
DEFINESCORE

- DEFINEMERGE
- MINUS
- WEIGHT
- ACCUM

For example, the following statement is valid:

```
DEFINESCORE('dog', OCCURRENCE) AND DEFINESCORE('cat', RELEVANCE)
```

Queries for the term "dog" using occurrence scoring, and the term "cat" using relevance scoring.

- If DEFINESCORE is used as a parameter of other operators, then an error will be returned. For example, the following example returns an error:

```
SYN(DEFINESCORE('cat', OCCURRENCE))
```

- When used with query templates, the scoring_expression overrides the values specified by the template. For example,

```
query>
  <textquery grammar="CONTEXT" lang="english">
    DEFINESCORE(dog, RELEVANCE) and cat
  </textquery>
  <score datatype="INTEGER" algorithm="COUNT"/>
</query>
```

Queries for "dog" and "cat", scores "cat" using OCCURRENCE(COUNT), and scores "dog" based on RELEVANCE.

**Related Topic**

- DEFINEMERGE on page 3-12.
**EQUIValence (=)**

Use the EQUIV operator to specify an acceptable substitution for a word in a query.

**Syntax**

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>_term1=<em>term2</em></td>
<td>Specifies that <em>term2</em> is an acceptable substitution for <em>term1</em>. Score calculated as the sum of all occurrences of both terms.</td>
</tr>
<tr>
<td><em>term1 equiv term2</em></td>
<td></td>
</tr>
</tbody>
</table>

**Example**

The following example returns all documents that contain either the phrase *alsatians are big dogs* or *labradors are big dogs*:

'labradors=alsatians are big dogs'

**Operator Precedence**

The EQUIV operator has higher precedence than all other operators except the expansion operators (fuzzy, soundex, stem).
Fuzzy

Use the fuzzy operator to expand queries to include words that are spelled similarly to the specified term. This type of expansion is helpful for finding more accurate results when there are frequent misspellings in your document set.

The fuzzy syntax enables you to rank the result set so that documents that contain words with high similarity to the query word are scored higher than documents with lower similarity. You can also limit the number of expanded terms.

Unlike stem expansion, the number of words generated by a fuzzy expansion depends on what is in the index. Results can vary significantly according to the contents of the index.

Supported Languages

Oracle Text supports fuzzy definitions for English, French, German, Italian, Dutch, Spanish, Portuguese, Japanese, OCR, and auto-language detection.

Stopwords

If the fuzzy expansion returns a stopword, the stopword is not included in the query or highlighted by CTX_DOC.HIGHLIGHT or CTX_DOC.MARKUP.

Base-Letter Conversion

If base-letter conversion is enabled for a text column and the query expression contains a fuzzy operator, Oracle Text operates on the base-letter form of the query.

Syntax

fuzzy(term, score, numresults, weight)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term</td>
<td>Specify the word on which to perform the fuzzy expansion. Oracle Text expands term to include words only in the index. The word needs to be at least 3 characters for the fuzzy operator to process it.</td>
</tr>
<tr>
<td>score</td>
<td>Specify a similarity score. Terms in the expansion that score below this number are discarded. Use a number between 1 and 80. The default is 60.</td>
</tr>
<tr>
<td>numresults</td>
<td>Specify the maximum number of terms to use in the expansion of term. Use a number between 1 and 5000. The default is 100.</td>
</tr>
<tr>
<td>weight</td>
<td>Specify WEIGHT or W for the results to be weighted according to their similarity scores. Specify NOWEIGHT or N for no weighting of results.</td>
</tr>
</tbody>
</table>

Examples

Consider the CONTAINS query:

...CONTAINS(TEXT, 'fuzzy(government, 70, 6, weight)', 1) > 0;

This query expands to the first six fuzzy variations of government in the index that have a similarity score over 70.
In addition, documents in the result set are weighted according to their similarity to government. Documents containing words most similar to government receive the highest score.

Skip unnecessary parameters using the appropriate number of commas. For example:

'fuzzy(government,,,,weight)'

**Backward Compatibility Syntax**

The old fuzzy syntax from previous releases is still supported. This syntax is as follows:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>?term</td>
<td>Expands term to include all terms with similar spellings as the specified term. Term needs to be at least 3 characters for the fuzzy operator to process it.</td>
</tr>
</tbody>
</table>
HASPATH

Use this operator to find all XML documents that contain a specified section path. You can also use this operator to do section equality testing.

Your index must be created with the PATH SECTION GROUP for this operator to work.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HASPATH(path)</td>
<td>Searches an XML document set and returns a score of 100 for all documents where path exists. Separate parent and child paths with the / character. For example, you can specify A/B/C.</td>
</tr>
<tr>
<td></td>
<td>See example.</td>
</tr>
<tr>
<td>HASPATH(A=&quot;value&quot;)</td>
<td>Searches an XML document set and returns a score of 100 for all documents that have the element A with content value and only value.</td>
</tr>
<tr>
<td></td>
<td>See example.</td>
</tr>
</tbody>
</table>

Using Special Characters with HASPATH and INPATH

The following rules govern the use of special characters with regard to both the HASPATH and INPATH operators:

- Left-brace (⟨) and right-brace (⟩) characters are not allowed inside HASPATH or INPATH expressions unless they are inside the equality operand enclosed by double quotes. So both ‘HASPATH (⟨ A/B ⟩)’ and ‘HASPATH (/A/⟨ B ⟩)’ will return errors. However, ‘HASPATH (/A[B="{author}"])’ will be parsed correctly.

- With exception of the backslash (\), special characters, such as dollar sign ($), percent sign (%), underscore (_), left brace ({), and right brace (}), when inside the equality operand enclosed by double or single quotes, have no special meaning. (That is, no stemming, wildcard expansion, or similar processing will be performed on them.) However, they are still subject to regular text lexing and will be translated to whitespace, with the exception of characters declared as printjoins. A backslash will still escape any character that immediately follows it.

For example, if the hyphen (-) and the double quote character (”) are defined as printjoins in a lexer preference, then:

- The string B_TEXT inside HASPATH (/A[B="B_TEXT"] ) will be lexed as the phrase B TEXT.
- The string B-TEXT inside HASPATH (/A[B="B-TEXT"] ) will be lexed as the word B-TEXT.
- The string B”TEXT inside HASPATH (/A[B="B ’TEXT"] ) will be lexed as the word B”TEXT. You must use a backslash to escape the double quote between B and TEXT, or you will get a parsing error.
- The string (B_TEXT) inside HASPATH (/A[B=" (B_TEXT) "] ) will be lexed as a phrase B TEXT.
**Example**

**Path Testing**
The query

\[
\text{HASPATH (A/B/C)}
\]

finds and returns a score of 100 for the document

\[
\text{<A><B><C>dog</C></B></A>}
\]

without the query having to reference *dog* at all.

**Section Equality Testing**
The query

\[
\text{dog INPATH A}
\]

finds

\[
\text{<A>dog</A>}
\]

but it also finds

\[
\text{<A>dog park</A>}
\]

To limit the query to the term *dog* and nothing else, you can use a section equality test with the HASTPATH operator. For example,

\[
\text{HASPATH (A=\text{'dog'})}
\]

finds and returns a score of 100 only for the first document, and not the second.

**Limitations**

Because of how XML section data is recorded, false matches might occur with XML sections that are completely empty as follows:

\[
\text{<A><B><C></C></B><D><E></E></D></A>}
\]

A query of \text{HASPATH (A/B/E)} or \text{HASPATH (A/D/C)} falsely matches this document. This type of false matching can be avoided by inserting text between empty tags.
INPATH

Use this operator to do path searching in XML documents. This operator is like the WITHIN operator except that the right-hand side is a parentheses enclosed path, rather than a single section name.

Your index must be created with the PATH_SECTION_GROUP for the INPATH operator to work.

Syntax

The INPATH operator has the following syntax:

**Top-Level Tag Searching**

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (/A)</td>
<td>Returns documents that have <em>term</em> within the &lt;A&gt; and &lt;/A&gt; tags.</td>
</tr>
<tr>
<td>term INPATH (A)</td>
<td></td>
</tr>
</tbody>
</table>

**Any-Level Tag Searching**

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (//A)</td>
<td>Returns documents that have <em>term</em> in the &lt;A&gt; tag at any level. This query is the same as 'term WITHIN A'.</td>
</tr>
</tbody>
</table>

**Direct Parentage Path Searching**

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (A/B)</td>
<td>Returns documents where <em>term</em> appears in a B element which is a direct child of a top-level A element. For example, a document containing &lt;A&gt;&lt;B&gt;*term&lt;/B&gt;&lt;/A&gt; is returned.</td>
</tr>
</tbody>
</table>

**Single-Level Wildcard Searching**

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (A/*/B)</td>
<td>Returns documents where <em>term</em> appears in a B element which is a grandchild (two levels down) of a top-level A element. For example a document containing &lt;A&gt;&lt;D&gt;&lt;B&gt;*term&lt;/B&gt;&lt;/D&gt;&lt;/A&gt; is returned.</td>
</tr>
</tbody>
</table>
Multi-level Wildcard Searching

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (A/<em>/B/</em>/C)</td>
<td>Returns documents where term appears in a C element which is 3 levels down from a B element which is two levels down (grandchild) of a top-level A element.</td>
</tr>
</tbody>
</table>

Any-Level Descendant Searching

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (A//B)</td>
<td>Returns documents where term appears in a B element which is some descendant (any level) of a top-level A element.</td>
</tr>
</tbody>
</table>

Attribute Searching

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (//A/@B)</td>
<td>Returns documents where term appears in the B attribute of an A element at any level. Attributes must be bound to a direct parent.</td>
</tr>
</tbody>
</table>

Descendant/Attribute Existence Testing

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (A[B])</td>
<td>Returns documents where term appears in a top-level A element which has a B element as a direct child.</td>
</tr>
<tr>
<td>term INPATH (A[../B])</td>
<td>Returns documents where term appears in a top-level A element which has a B element as a descendant at any level.</td>
</tr>
<tr>
<td>term INPATH (//A[@B])</td>
<td>Finds documents where term appears in an A element at any level which has a B attribute. Attributes must be tied to a direct parent.</td>
</tr>
</tbody>
</table>

Attribute Value Testing

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (A[@B = &quot;value&quot;]))</td>
<td>Finds all documents where term appears in a top-level A element which has a B attribute whose value is value.</td>
</tr>
<tr>
<td>term INPATH (A[@B != &quot;value&quot;]))</td>
<td>Finds all documents where term appears in a top-level A element which has a B attribute whose value is not value.</td>
</tr>
</tbody>
</table>

Tag Value Testing

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (A[B = &quot;value&quot;]))</td>
<td>Returns documents where term appears in an A tag which has a B tag whose value is value.</td>
</tr>
</tbody>
</table>
Not

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (A[NOT(B)])</td>
<td>Finds documents where term appears in a top-level A element which does not have a B element as an immediate child.</td>
</tr>
</tbody>
</table>

AND and OR Testing

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (A[B and C])</td>
<td>Finds documents where term appears in a top-level A element which has a B and a C element as an immediate child.</td>
</tr>
<tr>
<td>term INPATH (A[B and @C=&quot;value&quot;]])</td>
<td>Finds documents where term appears in a top-level A element which has a B element and a C attribute whose value is value.</td>
</tr>
<tr>
<td>term INPATH (A[B OR C])</td>
<td>Finds documents where term appears in a top-level A element which has a B element or a C element.</td>
</tr>
</tbody>
</table>

Combining Path and Node Tests

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term INPATH (A[@B = &quot;value&quot;]/C/D)</td>
<td>Returns documents where term appears in aD element which is the child of a C element, which is the child of a top-level A element with a B attribute whose value is value.</td>
</tr>
</tbody>
</table>

Nested INPATH

Nest the entire INPATH expression in another INPATH expression as follows:

(dog INPATH (///A/B/C)) INPATH (D)

When you do so, the two INPATH paths are completely independent. The outer INPATH path does not change the context node of the inner INPATH path. For example:

(dog INPATH (A)) INPATH (D)

never finds any documents, because the inner INPATH is looking for dog within the top-level tag A, and the outer INPATH constrains that to document with top-level tag D. A document can have only one top-level tag, so this expression never finds any documents.

Case-Sensitivity

Tags and attribute names in path searching are case-sensitive. That is,

dog INPATH (A)

finds <A>dog</A> but does not find <a>dog</a>. Instead use

dog INPATH (a)
Using Special Characters with INPATH

See "Using Special Characters with HASPATH and INPATH" on page 3-20 for information on using special characters, such as the percent sign (%) or the backslash (\), with INPATH.

Examples

**Top-Level Tag Searching**
To find all documents that contain the term *dog* in the top-level tag `<A>`:

```
dog INPATH (/A)
```

or

```
dog INPATH(A)
```

**Any-Level Tag Searching**
To find all documents that contain the term *dog* in the `<A>` tag at any level:

```
dog INPATH(//A)
```

This query finds the following documents:

```
<A>dog</A>
```

and

```
<C><B><A>dog</A></B></C>
```

**Direct Parentage Searching**
To find all documents that contain the term *dog* in a B element that is a direct child of a top-level A element:

```
dog INPATH(A/B)
```

This query finds the following XML document:

```
<A><B>My dog is friendly.</B><A>
```

but does not find:

```
<C><B>My dog is friendly.</B></C>
```

**Tag Value Testing**
You can test the value of tags. For example, the query:

```
dog INPATH(A[@B='dog'])
```

Finds the following document:

```
<A><B>dog</B></A>
```

But does not find:

```
<A><B>My dog is friendly.</B></A>
```

**Attribute Searching**
You can search the content of attributes. For example, the query:

```
dog INPATH(//A/@B)
```
Finds the document

```
<C><A B="snoop dog"> </A> </C>
```

**Attribute Value Testing**

You can test the value of attributes. For example, the query

```xml
California INPATH (///A[@B = 'home address'])
```

Finds the document:

```
<A B="home address">San Francisco, California, USA</A>
```

But does not find:

```
<A B="work address">San Francisco, California, USA</A>
```

**Path Testing**

You can test if a path exists with the `HASPATH` operator. For example, the query:

```
HASPATH(A/B/C)
```

finds and returns a score of 100 for the document

```
<A><B><C>dog</C></B></A>
```

without the query having to reference `dog` at all.

**Limitations**

**Testing for Equality**

The following is an example of an `INPATH` equality test.

```xml
dog INPATH (A[@B = 'foo'])
```

The following limitations apply for these expressions:

- Only equality and inequality are supported. Range operators and functions are not supported.
- The left hand side of the equality must be an attribute. Tags and literals here are not enabled.
- The right hand side of the equality must be a literal. Tags and attributes here are not allowed.
- The test for equality depends on your lexer settings. With the default settings, the query
  ```xml
dog INPATH (A[@B= "pot of gold"])```  
  matches the following sections:
  ```xml
  <A B="POT OF GOLD">dog</A>
  ```
  and
  ```xml
  <A B="pot of gold">dog</A>
  ```
  because lexer is case-insensitive by default.

```
<A B="POT IS GOLD">dog</A>
```
because *of* and *is* are default stopwords in English, and a stopword matches any stopword word.

```xml
<A B="POT_OF_GOLD">dog</A>
```

because the underscore character is not a join character by default.
Use the `MDATA` operator to query documents that contain `MDATA` sections. `MDATA` sections are metadata that have been added to documents to speed up mixed querying.

`MDATA` queries are treated exactly as literals. For example, with the query:

```
MDATA(price, $1.24)
```

the `$` is not interpreted as a stem operator, nor is the `. (period) transformed into whitespace. A right (close) parenthesis terminates the `MDATA` operator, so that `MDATA` values that have close parentheses cannot be searched.

### Syntax

```
MDATA(sectionname, value)
```

<table>
<thead>
<tr>
<th>sectionname</th>
<th>The name of the <code>MDATA</code> section(s) to search. <code>MDATA</code> will also search <code>DATE</code> or numerical equality if the <code>sectionname</code> parameter is mapped to a <code>FILTER BY</code> column of <code>DATE</code> or some numerical type.</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
<td>The value of the <code>MDATA</code> section. For example, if an <code>MDATA</code> section called <code>Booktype</code> has been created, it might have a value of <code>paperback</code>. For <code>MDATA</code> operator on <code>MDATA</code> sections that are mapped to a <code>DATE FILTER BY</code> column, the <code>MDATA</code> value must follow the Date format: <code>YYYY-MM-DD HH24:MI:SS</code>. Otherwise, the expected rows will not be returned. If the time component is omitted, it will default to <code>00:00:00</code>, according to SQL semantics.</td>
</tr>
</tbody>
</table>

### Example

Suppose you want to query for books written by the writer `Nigella Lawson` that contain the word `summer`. Assuming that an `MDATA` section called `AUTHOR` has been declared, you can query as follows:

```
SELECT id FROM idx_docs
WHERE CONTAINS(text, 'summer AND MDATA(author, Nigella Lawson)')>0
```

This query will only be successful if an `AUTHOR` tag has the exact value `Nigella Lawson` (after simplified tokenization). `Nigella` or `Ms. Nigella Lawson` will not work.

### Notes

- `MDATA` query values ignore stopwords.
- The `MDATA` operator returns 100 or 0, depending on whether the document is a match.
- The `MDATA` operator is not supported for `CTXCAT`, `CTXRULE`, or `CTXXPATH` indexes.
- Table 3–2 shows how `MDATA` interacts with some other query operators:
When MDATA sections repeat, each instance is a separate and independent value. For instance, the document

<AUTHOR>Terry Pratchett</AUTHOR><AUTHOR>Douglas Adams</AUTHOR>

can be found with any of the following queries:

MDATA(author, Terry Pratchett)
MDATA(author, Douglas Adams)
MDATA(author, Terry Pratchett) and MDATA(author, Douglas Adams)

but not any of the following:

MDATA(author, Terry Pratchett Douglas Adams)
MDATA(author, Terry Pratchett & Douglas Adams)
MDATA(author, Pratchett Douglas)

### Related Topics

See also "ADD_MDATA" on page 7-9 and "ADD_MDATA_SECTION" on page 7-12, as well as the Section Searching chapter of the Oracle Text Application Developer's Guide.
MINUS (-)

Use the MINUS operator to lower the score of documents that contain unwanted noise terms. MINUS is useful when you want to search for documents that contain one query term but want the presence of a second term to cause a document to be ranked lower.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term1-term2</td>
<td>Returns documents that contain term1. Calculates score by subtracting the score of term2 from the score of term1. Only documents with positive score are returned.</td>
</tr>
<tr>
<td>term1 minus term2</td>
<td></td>
</tr>
</tbody>
</table>

Example

Suppose a query on the term cars always returned high scoring documents about Ford cars. You can lower the scoring of the Ford documents by using the expression:

'cars - Ford'

In essence, this expression returns documents that contain the term cars and possibly Ford. However, the score for a returned document is the score of cars minus the score of Ford.

Related Topics

See Also: "NOT (¬)" on page 3-40
The Mild Not (MNOT) operator is similar to the NOT and MINUS operators. The Mild Not operator returns hits where the the left child is not contained by the right child. Both children can only be TERM or PHRASE nodes.

The semantics can be illustrated with a query of "term1 mnot term1 term2", where the hits for "term1 term2" will be filtered out. For example:

- A document with only term1 will be returned, with score unchanged.
- A document with only term1 term2 will not be returned.
- A document with term1 term1 term2 will be returned, but the score will be calculated using just the first term1 hit.

The behavior described in the third bullet is different from the behavior of NOT, which does not return this type of document.

The MNOT operator is more specific than the MINUS operator, in that the left child must be contained by the right child. If it is not, the Mild Not operator ignores the right child. Also, for Mild Not, the right child is a true filter, that is, it does not simply subtract the scores of left child and right child.

The MNOT operator has precedence lower than NOT and higher than WITHIN.

### Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>term1 mnot term1 term2</code></td>
<td>Returns docs that contain term1 unless it is part of the phrase term1 term2.</td>
</tr>
<tr>
<td><code>term1 mnot term2</code></td>
<td>Returns all documents that contain term1. It will be the same query as just term1.</td>
</tr>
</tbody>
</table>

### Example

The children of the MNOT operator must be a TERM or PHRASE.

```sql
SELECT * FROM docs
WHERE CONTAINS(txt, 'term1 mnot term1 term2') >0
```

### Related Topics

See Also: "NOT (−)" on page 3-40
Narrower Term (NT, NTG, NTP, NTI)

Use the narrower term operators (NT, NTG, NTP, NTI) to expand a query to include all the terms that have been defined in a thesaurus as the narrower or lower level terms for a specified term. They can also expand the query to include all of the narrower terms for each narrower term, and so on down through the thesaurus hierarchy.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NT(\text{term})[\text{level}][\text{thes}]</td>
<td>Expands a query to include all the lower level terms defined in the thesaurus as narrower terms for \text{term}.</td>
</tr>
<tr>
<td>NTG(\text{term})[\text{level}][\text{thes}]</td>
<td>Expands a query to include all the lower level terms defined in the thesaurus as narrower generic terms for \text{term}.</td>
</tr>
<tr>
<td>NTP(\text{term})[\text{level}][\text{thes}]</td>
<td>Expands a query to include all the lower level terms defined in the thesaurus as narrower partitive terms for \text{term}.</td>
</tr>
<tr>
<td>NTI(\text{term})[\text{level}][\text{thes}]</td>
<td>Expands a query to include all the lower level terms defined in the thesaurus as narrower instance terms for \text{term}.</td>
</tr>
</tbody>
</table>

\text{term}
Specify the operand for the narrower term operator. \text{term} is expanded to include the narrower term entries defined for the term in the thesaurus specified by \text{thes}. The number of narrower terms included in the expansion is determined by the value for \text{level}. You cannot specify expansion operators in the \text{term} argument.

\text{qualifier}
Specify a qualifier for \text{term}, if \text{term} is a homograph (word or phrase with multiple meanings, but the same spelling) that appears in two or more nodes in the same hierarchy branch of \text{thes}.

If a qualifier is not specified for a homograph in a narrower term query, the query expands to include all of the narrower terms of all homographic terms.

\text{level}
Specify the number of levels traversed in the thesaurus hierarchy to return the narrower terms for the specified term. For example, a level of 1 in an \text{NT} query returns all the narrower term entries, if any exist, for the specified term. A level of 2 returns all the narrower term entries for the specified term, as well as all the narrower term entries, if any exist, for each narrower term.

The level argument is optional and has a default value of one (1). Zero or negative values for the level argument return only the original query term.

\text{thes}
Specify the name of the thesaurus used to return the expansions for the specified term. The \text{thes} argument is optional and has a default value of DEFAULT. A thesaurus named DEFAULT must exist in the thesaurus tables if you use this default value.

\textbf{Note}: If you specify \text{thes}, then you must also specify \text{level}.
Examples

The following query returns all documents that contain either the term *cat* or any of the *NT* terms defined for *cat* in the DEFAULT thesaurus:

```
'NT(cat)'
```

If you specify a thesaurus name, then you must also specify *level* as in:

```
'NT(cat, 2, mythes)'
```

The following query returns all documents that contain either *fairy tale* or any of the narrower instance terms for *fairy tale* as defined in the DEFAULT thesaurus:

```
'NTI(fairy tale)'
```

That is, if the terms *cinderella* and *snow white* are defined as narrower term instances for *fairy tale*, Oracle Text returns documents that contain *fairy tale*, *cinderella*, or *snow white*.

Notes

Each hierarchy in a thesaurus represents a distinct, separate branch, corresponding to the four narrower term operators. In a narrower term query, Oracle Text only expands the query using the branch corresponding to the specified narrower term operator.

Related Topics

Browse a thesaurus using procedures in the CTX_THES package.

**See Also:** CTX_THES.NT in Chapter 13, "CTX_THES Package" for more information on browsing the narrower terms in your thesaurus
NDATA

Use the NDATA operator to find matches that are spelled in a similar way or where rearranging the terms of the specified phrase is useful. It is helpful for finding more accurate results when there are frequent misspellings (or inaccurate orderings) of name data in the document set. This operator can be used only on defined NDATA sections. The NDATA syntax enables you to rank the result set so that documents that contain words with high orthographic similarity are scored higher than documents with lower similarity.

Normalization

A lexer does not process NDATA query phrases. Users, however, set base letter and alternate spelling attributes for a particular section group containing NDATA sections. Query case is normalized and non-character data (except for white space) is removed (for example, numerical or punctuation).

Syntax

ndata(sectionname, phrase [,order][,proximity])

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Default Value</th>
<th>Parameter Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sectionname</td>
<td>Specify the name of a defined NDATA sections to query (that is, section_name)</td>
<td></td>
</tr>
<tr>
<td>phrase</td>
<td>Specify the phrase for the name data query. The phrase parameter can be a single word or a phrase, or a string of words in free text format. The score returned is a relevant score. Oracle Text ignores any query operators that are included in phrase. The phrase should be a minimum of two characters in length and should not exceed 4000 characters in length.</td>
<td></td>
</tr>
<tr>
<td>order</td>
<td>NOORDER</td>
<td>Specify whether individual tokens (terms) in a query should be matched in-order or in any order. The order parameter provides a primary filter for matching candidate documents. ORDER or O - The query terms are matched in-order. NOORDER or N [DEFAULT] - The query terms are matched in any order.</td>
</tr>
<tr>
<td>proximity</td>
<td>NOPROXIMITY</td>
<td>Specify whether the proximity of terms should influence the similarity score of candidate matches. That is, if the proximity parameter is enabled, non-matching additional terms between matching terms will reduce the similarity score of candidate matches. PROXIMITY or P - The similarity score influenced by the proximity of query terms in candidate matches. NOPROXIMITY or N [DEFAULT] - The similarity score is not influenced by the proximity of query terms in candidate matches.</td>
</tr>
</tbody>
</table>
Examples

An NDATA query on an indexed surname section name that matches terms in the query phrase in any order without influencing the similarity score by the proximity of the black and smith terms has the form:

```
SELECT entryid, SCORE(1) FROM people WHERE
CONTAINS(idx_column, 'NDATA(surname, black smith)',1)>0;
```

An NDATA query on an indexed surname section name that matches terms in the query phrase in any order and in which similarity scores are influenced by the proximity of the black and smith terms has the form:

```
SELECT entryid, SCORE(1) FROM people WHERE
CONTAINS(idx_column, 'NDATA(surname, black smith,proximity)',1)>0;
```

An NDATA query on an indexed surname section name that matches terms in the query phrase in-order without influencing the similarity score by the proximity of the black and smith terms has the form:

```
SELECT entryid, SCORE(1) FROM people WHERE
CONTAINS(idx_column, 'NDATA(surname, black smith,order)',1)>0;
```

An NDATA query on an indexed surname section name that matches terms in the query phrase in-order and in which similarity scores are influenced by the proximity of the black and smith terms has the form:

```
SELECT entryid, SCORE(1) FROM people WHERE
CONTAINS(idx_column, 'NDATA(surname, black smith,order,proximity)',1)>0;
```

Notes

The NDATA query operator does not provide offset information. As such, it cannot be used as a child of WITHIN, NEAR();, or EQUIV(=), and NDATA sections will be ignored by CTX_DOC.HIGHLIGHT, CTX_DOC.SNIPPET, and CTX_DOC.MARKUP. The NDATA operator also is not supported in the CTXCAT grammar. It can be used with other operators, including OR and query templates.

A use case of the NDATA operator may involve finding a particular entry based on an approximate spelling of a person’s full-name and an estimated date-of-birth. Supposing the entries’ date-of-births are stored as an SDATA section, user-defined scoring’s alternate scoring template can be used to combine the scores of the full-name’s NDATA section data and the date-of-birth’s SDATA section data.

The name john smith is queried for the section specified by the fullname section_name. Altering the NDATA operator’s score based on the closeness of the SDATA section’s date-of-birth to the date 08-NOV-2005 modifies the ranking of matching documents:

```
<query>
  <textquery grammar="CONTEXT" lang="english">
    NDATA(fullname, john smith)
  </textquery>
  <score algorithm="COUNT" normalization_expr =
  "doc_score-(DATE(8-NOV-2005)-sdata:dob)/"/>
</query>
```
Use the `NEAR` operator to return a score based on the proximity of two or more query terms. Oracle Text returns higher scores for terms closer together and lower scores for terms farther apart in a document.

**Note:** The `NEAR` operator works with only word queries. You cannot use `NEAR` in `ABOUT` queries.

### Syntax

**word 1-n**
Specify the terms in the query separated by commas. The query terms can be single words or phrases and may make use of other query operators (see "NEAR with Other Operators").

**max_span**
Optionally specify the size of the biggest clump. The default is 100. Oracle Text returns an error if you specify a number greater than 100.

A clump is the smallest group of words in which all query terms occur. All clumps begin and end with a query term.

For near queries with two terms, `max_span` is the maximum distance allowed between the two terms. For example, to query on `dog` and `cat` where `dog` is within 6 words of `cat`, enter the following query:

`'near((dog, cat), 6)'`

**order**
Specify `TRUE` for Oracle Text to search for terms in the order you specify. The default is `FALSE`.

For example, to search for the words `monday`, `tuesday`, and `wednesday` in that order with a maximum clump size of 20, enter the following query:

`'near((monday, tuesday, wednesday), 20, TRUE)'`

**Note:** To specify `order`, then you must always specify a number for `max_span`.

Oracle Text might return different scores for the same document when you use identical query expressions that have the `order` flag set differently. For example, Oracle Text might return different scores for the same document when you enter the following queries:

`'near((dog, cat), 50, FALSE)'`
`'near((dog, cat), 50, TRUE)'`
NEAR Scoring

The scoring for the NEAR operator combines frequency of the terms with proximity of terms. For each document that satisfies the query, Oracle Text returns a score between 1 and 100 that is proportional to the number of clumps in the document and inversely proportional to the average size of the clumps. This means many small clumps in a document result in higher scores, because small clumps imply closeness of terms.

The number of terms in a query also affects score. Queries with many terms, such as seven, generally need fewer clumps in a document to score 100 than do queries with few terms, such as two.

A clump is the smallest group of words in which all query terms occur. All clumps begin and end with a query term. Define clump size with the max_span parameter, as described in this section.

The size of a clump does not include the query terms themselves. So for the query NEAR((DOG, CAT), 1), dog cat will be a match, and dog ate cat will be a match, but dog sat on cat will not be a match.

NEAR with Other Operators

You can use the NEAR operator with other operators such as AND and OR. Scores are calculated in the regular way.

For example, to find all documents that contain the terms tiger, lion, and cheetah where the terms lion and tiger are within 10 words of each other, enter the following query:

'near((lion, tiger), 10) AND cheetah'

The score returned for each document is the lower score of the near operator and the term cheetah.

You can also use the equivalence operator to substitute a single term in a near query:

'near((stock crash, Japan=Korea), 20)'

This query asks for all documents that contain the phrase stock crash within twenty words of Japan or Korea.

The following NEAR syntax is now valid:

```
SELECT * FROM docs WHERE CONTAINS(txt, 'near((aterm1 aterm2 ... atermI
OR bterm1 bterm2 ... btermJ
OR cterm1 cterm2 ... ctermK, dterm))') >0
```

There can be any number of ORs in a given NEAR child, and the OR can appear in any of the NEAR children.

The NEAR within NEAR feature allows users to use nested proximity queries. Users can execute queries such as the following:

```
SELECT * FROM docs
WHERE CONTAINS(txt, 'near((near((term1, term2),5), term3), 100)')>0
```

This will return documents where term1, term2, and term3 are near within a 100 token window and, additionally, the tokens term1 and term2 are near within a 5 token window.

Mixing the semicolon and NEAR syntax is not supported and will throw an error. That is, the queries "near((a;b,c), 3)" or "near((a,b));c" will be disallowed.

The following operators also work with NEAR and ;:
- EQUIV
- All expansion operators that produce words, phrases, or EQUIV. These include:
  - soundex
  - fuzzy
  - wildcards
  - stem

**Backward Compatibility NEAR Syntax**

You can write near queries using the syntax of previous Oracle Text releases. For example, to find all documents where lion occurs near tiger, write:

'lion near tiger'

or with the semi-colon as follows:

'lion;tiger'

This query is equivalent to the following query:

'near((lion, tiger), 100, FALSE)'

---

**Note:** Only the syntax of the NEAR operator is backward compatible. In the example, the score returned is calculated using the clump method as described in this section.

---

**Highlighting with the NEAR Operator**

When you use highlighting and your query contains the near operator, all occurrences of all terms in the query that satisfy the proximity requirements are highlighted. Highlighted terms can be single words or phrases.

For example, assume a document contains the following text:

Chocolate and vanilla are my favorite ice cream flavors. I like chocolate served in a waffle cone, and vanilla served in a cup with carmel syrup.

If the query is near((chocolate, vanilla)), 100, FALSE), the following is highlighted:

<<Chocolate>> and <<vanilla>> are my favorite ice cream flavors. I like chocolate served in a waffle cone, and vanilla served in a cup with carmel syrup.

However, if the query is near((chocolate, vanilla)), 4, FALSE), only the following is highlighted:

<<Chocolate>> and <<vanilla>> are my favorite ice cream flavors. I like chocolate served in a waffle cone, and vanilla served in a cup with carmel syrup.

**See Also:** Chapter 8, "CTX_DOC Package" for more information about the procedures for highlighting

**Section Searching and NEAR**

Use the NEAR operator with the WITHIN operator for section searching as follows:

'near((dog, cat), 10) WITHIN Headings'
When evaluating expressions such as these, Oracle Text looks for clumps that lie entirely within the given section.

In this example, only those clumps that contain *dog* and *cat* that lie entirely within the section *Headings* are counted. That is, if the term *dog* lies within *Headings* and the term *cat* lies five words from *dog*, but outside of *Headings*, this pair of words does not satisfy the expression and is not counted.
NOT (~)

Use the NOT operator to search for documents that contain one query term and not another.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term1~term2</td>
<td>Returns documents that contain term1 and not term2.</td>
</tr>
<tr>
<td>term1 not term2</td>
<td></td>
</tr>
</tbody>
</table>

Examples

To obtain the documents that contain the term *animals* but not *dogs*, use the following expression:

'animals ~ dogs'

Similarly, to obtain the documents that contain the term *transportation* but not *automobiles* or *trains*, use the following expression:

'transportation not (automobiles or trains)'

Note: The NOT operator does not affect the scoring produced by the other logical operators.

Related Topics

See Also: "MINUS (-)" on page 3-30
OR (I)

Use the OR operator to search for documents that contain at least one occurrence of any of the query terms.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>term1</td>
<td>term2</td>
</tr>
<tr>
<td>term1</td>
<td>or</td>
</tr>
</tbody>
</table>

Examples

To obtain the documents that contain the term cats or the term dogs, use either of the following expressions:

'cats | dogs'
'cats OR dogs'

Scoring

In an OR query, the score returned is the score for the highest query term. In the example, if the scores for cats and dogs is 30 and 40 within a document, the document scores 40.

Related Topics

See Also: The OR operator returns documents that contain any of the query terms, while the AND operator returns documents that contain all query terms. See “AND (&)” on page 3-9.
Preferred Term (PT)

Use the preferred term operator (PT) to replace a term in a query with the preferred term that has been defined in a thesaurus for the term.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT(term[,thes])</td>
<td>Replaces the specified word in a query with the preferred term for term.</td>
</tr>
</tbody>
</table>

**term**
Specify the operand for the preferred term operator. *term* is replaced by the preferred term defined for the term in the specified thesaurus. However, if no PT entries are defined for the term, *term* is not replaced in the query expression and *term* is the result of the expansion.

You cannot specify expansion operators in the *term* argument.

**thes**
Specify the name of the thesaurus used to return the expansions for the specified term. The *thes* argument is optional and has a default value of DEFAULT. As a result, a thesaurus named DEFAULT must exist in the thesaurus tables before using any of the thesaurus operators.

Example

The term *automobile* has a preferred term of *car* in a thesaurus. A PT query for *automobile* returns all documents that contain the word *car*. Documents that contain the word *automobile* are not returned.

Related Topics

Browse a thesaurus using procedures in the CTX_THES package.

See Also: CTX_THES.PT in Chapter 13, "CTX_THES Package" for more information on browsing the preferred terms in your thesaurus.
Related Term (RT)

Use the related term operator (RT) to expand a query to include all related terms that have been defined in a thesaurus for the term.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT(term[,thes])</td>
<td>Expands a query to include all the terms defined in the thesaurus as a related term for term.</td>
</tr>
</tbody>
</table>

**term**
Specify the operand for the related term operator. term is expanded to include term and all the related entries defined for term in thes.

You cannot specify expansion operators in the term argument.

**thes**
Specify the name of the thesaurus used to return the expansions for the specified term. The thes argument is optional and has a default value of DEFAULT. As a result, a thesaurus named DEFAULT must exist in the thesaurus tables before using any of the thesaurus operators.

Example

The term dog has a related term of wolf. An RT query for dog returns all documents that contain the word dog and wolf.

Related Topics

Browse a thesaurus using procedures in the CTX_THES package.

**See Also:** CTX_THES.RT in Chapter 13, "CTX_THES Package" for more information on browsing the related terms in your thesaurus.
**SDATA**

Use the `SDATA` operator to perform tests on `SDATA` sections and columns, which contain structured data values. `SDATA` sections speed up mixed querying and ordering. This operator provides structured predicate support for `CONTAINS`, which extends non-SQL interfaces such as `count_hits` or the result set interface.

`SDATA` operators should only be used as descendants of `AND` operators that also have non-`SDATA` children.

`SDATA` queries perform on string or numeric literals, and on date strings. The string literal and date string are enclosed within single or double quote characters. The numeric value is not enclosed in quote characters, and must conform to the SQL format of `NUMBER`. For example:

```sql
CONTAINS(text, 'dog and SDATA(category = 'news')')>0 ...
SDATA(rating between 1.2 and 3.4) ...
SDATA(author LIKE 'FFORDE%') ...
SDATA(date >= '2005-09-18') ...
```

Closed parentheses are permitted, as long as they are enclosed in single or double quotes.

The `SDATA` operator can be used in query templates.

**Syntax**

```
Syntax
SData ::= "SDATA" "(" SDataPredicate ")"
SDataPredicate ::= sectionname SDataTest
SDataTest ::= SDataSingleOp SDataLiteral | SDataBetweenOp | <"is" ("not")? "null">
SDataSingleOp ::= (<"" | "<=" | "=" | "<" | ">=" | ">" | "=" | ">" | "like") SDataLiteral
SDataBetweenOp ::= "between" SDataLiteral "and" SDataLiteral
SDataLiteral ::= numeric_literal | "" string_literal "" | "" date_string ""
```

**sectionname**
The name of the `SDATA` section(s) on which to search and perform the test, or check.

**SDataLiteral**
The value of the `SDATA` section. This must be either a string literal, numeric literal, or a date string.

The `SDATA` operator returns a score of 100 if the enclosed predicate returns `TRUE`, and returns 0 otherwise. In the case of a `NULL` value, the `SDATA` operator returns a score of 0 (since in SQL it would not return `TRUE`).

Multi-valued semantics are not defined, as multi-valued `SDATA` sections are not supported.

Comparison of strings is case sensitive. The `BINARY` collation is always used.
Examples

Suppose that you want to query for books in the fiction category that contain the word summer. Assuming that an SDATA section called CATEGORY has been declared, you can query as follows:

```sql
SELECT id FROM idx_docs
WHERE CONTAINS(text, 'summer AND SDATA(category = "fiction")')>0
```

Restrictions

- An error is raised if the section name is not a defined SDATA section. The source of the section (for example, tag versus column) is not important.
- The syntax precludes RHS SDATA and expressions.
- SDATA operators cannot be children of WITHIN, INPATH, HASPATH, or NEAR.
- The datatype of the named SDATA section must be compatible with the literal provided (and the operator, for example, LIKE) or an error is raised.
- SDATA operators are not supported in CTXRULE query documents.
- SDATA operators have no effect on highlighting.

Notes

Oracle recommends using SDATA operators only as descendants of AND operators that also have non-SDATA children. Essentially, use SDATA operators as secondary (that is, checking or non-driving) criteria. For instance, "find documents with DOG that also have price > 5", rather than "find documents with rating > 4". Other usage may operate properly, but may not have optimal performance.

The following examples are consistent with recommended use:

```sql
dog & SDATA(foo = 5)
```

The SDATA is a child of an AND operator that also has non-SDATA children.

```sql
dog & (SDATA(foo = 5) | SDATA(x = 1))
```

Although the SDATA operators here are children of OR, they are still descendants of an AND operator with non-SDATA children.

The following examples show use that is not recommended:

```sql
SDATA(foo = 5)
```

Here, SDATA is the only criteria and, therefore, the driving criteria.

```sql
dog | SDATA(bar = 9)
```

The SDATA in this example is a child of an OR operator rather than an AND.
$\text{SDATA}(\text{foo} = 5) \land \text{SDATA}(\text{bar} = 7)$

While both $\text{SDATA}$ operators in this example are descendants of $\text{AND}$, this $\text{AND}$ operator does not have non-$\text{SDATA}$ children.

**Related Topics**

- [ADD_SDATA_COLUMN on page 7-14](#)
- [ADD_SDATA_SECTION on page 7-16](#)
- [CTX_SECTIONS on page G-10 in Appendix G, "Oracle Text Views"](#)

**See Also:**

- [Oracle Database SQL Language Reference](#)
- Chapter 8, "Searching Document Sections in Oracle Text" in Oracle Text Application Developer's Guide
soundex (!)

Use the soundex (!) operator to expand queries to include words that have similar sounds; that is, words that sound like other words. This function enables comparison of words that are spelled differently, but sound alike in English.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>!term</td>
<td>Expands a query to include all terms that sound the same as the specified term (English-language text only).</td>
</tr>
</tbody>
</table>

Example

```
SELECT ID, COMMENT FROM EMP_RESUME
WHERE CONTAINS (COMMENT, '!SMYTHE') > 0 ;
```

ID COMMENT
-- ------------
23 Smith is a hard worker who..

Language

Soundex works best for languages that use a 7-bit character set, such as English. It can be used, with lesser effectiveness, for languages that use an 8-bit character set, such as many Western European languages.

If you have base-letter conversion specified for a text column and the query expression contains a soundex operator, then Oracle Text operates on the base-letter form of the query.
stem ($)

Use the stem ($) operator to search for terms that have the same linguistic root as the query term.

If you use the BASIC_LEXER to index your language, stemming performance can be improved by using the index_stems attribute.

The Oracle Text stemmer, licensed from XSoft Division of Xerox Corporation, supports the following languages with the BASIC_LEXER: English, French, Spanish, Italian, German, and Dutch.

Japanese stemming is supported with the JAPANESE_LEXER.

Specify your stemming language with the BASIC_WORDLIST wordlist preference.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$term</td>
<td>Expands a query to include all terms having the same stem or root word as the specified term.</td>
</tr>
</tbody>
</table>

Examples

<table>
<thead>
<tr>
<th>Input</th>
<th>Expands To</th>
</tr>
</thead>
<tbody>
<tr>
<td>$scream</td>
<td>scream screaming screamed</td>
</tr>
<tr>
<td>$distinguish</td>
<td>distinguish distinguished distinguishes</td>
</tr>
<tr>
<td>$guitars</td>
<td>guitars guitar</td>
</tr>
<tr>
<td>$commit</td>
<td>commit committed</td>
</tr>
<tr>
<td>$cat</td>
<td>cat cats</td>
</tr>
<tr>
<td>$sing</td>
<td>sang sung sing</td>
</tr>
</tbody>
</table>

Behavior with Stopwords

If stem returns a word designated as a stopword, the stopword is not included in the query or highlighted by CTX_QUERY.HIGHLIGHT or CTX_QUERY.MARKUP.

Related Topics

See Also: For more information about enabling the stem operator with BASIC_LEXER, see "BASIC_LEXER" in Chapter 2, "Oracle Text Indexing Elements".
Stored Query Expression (SQE)

Use the SQE operator to call a stored query expression created with the `CTX_QUERY.STORE_SQE` procedure.

Stored query expressions can be used for creating predefined bins for organizing and categorizing documents or to perform iterative queries, in which an initial query is refined using one or more additional queries.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQE(SQE_name)</td>
<td>Returns the results for the stored query expression SQE_name.</td>
</tr>
</tbody>
</table>

Examples

To create an SQE named disasters, use `CTX_QUERY.STORE_SQE` as follows:

```sql
begin
ctx_query.store_sqe('disasters', 'hurricane or earthquake or blizzard');
end;
```

This stored query expression returns all documents that contain either hurricane, earthquake or blizzard.

This SQE can then be called within a query expression as follows:

```sql
SELECT SCORE(1), docid FROM news
WHERE CONTAINS(resume, 'sqe(disasters)', 1)> 0
ORDER BY SCORE(1);
```

Limitations

Up to 100 stored query expressions (SQEs) can be stored in a single Text query. If a Text query has more than 100 SQEs, including nested SQEs, then the query fails and error DRG-50949 is raised.
SYNonym (SYN)

Use the synonym operator (SYN) to expand a query to include all the terms that have been defined in a thesaurus as synonyms for the specified term.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYN(term[thes])</td>
<td>Expands a query to include all the terms defined in the thesaurus as synonyms for term.</td>
</tr>
</tbody>
</table>

**term**

Specify the operand for the synonym operator. term is expanded to include term and all the synonyms defined for term in thes.

You cannot specify expansion operators in the term argument.

**thes**

Specify the name of the thesaurus used to return the expansions for the specified term. The thes argument is optional and has a default value of DEFAULT. A thesaurus named DEFAULT must exist in the thesaurus tables if you use this default value.

Examples

The following query expression returns all documents that contain the term dog or any of the synonyms defined for dog in the DEFAULT thesaurus:

'SYN(dog)'

**Compound Phrases in Synonym Operator**

Expansion of compound phrases for a term in a synonym query are returned as AND conjunctives.

For example, the compound phrase temperature + measurement + instruments is defined in a thesaurus as a synonym for the term thermometer. In a synonym query for thermometer, the query is expanded to:

{thermometer} OR {{temperature}&{measurement}&{instruments}}

Related Topics

Browse your thesaurus using procedures in the CTX_THES package.

See Also: CTX_THES.SYN in Chapter 13, "CTX_THES Package" for more information on browsing the synonym terms in your thesaurus
threshold (>)

Use the threshold operator (>) in two ways:

- at the expression level
- at the query term level

The threshold operator at the expression level eliminates documents in the result set that score below a threshold number.

The threshold operator at the query term level selects a document based on how a term scores in the document.

**Syntax**

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>expression&gt;(n)</td>
<td>Returns only those documents in the result set that score above the threshold (n).</td>
</tr>
<tr>
<td>term&gt;(n)</td>
<td>Within an expression, returns documents that contain the query term with score of at least (n).</td>
</tr>
</tbody>
</table>

**Examples**

At the expression level, to search for documents that contain *relational databases* and to return only documents that score greater than 75, use the following expression:

'\(\text{relational databases} > 75\)'

At the query term level, to select documents that have at least a score of 30 for *lion* and contain *tiger*, use the following expression:

'\((\text{lion} > 30) \text{ and tiger}\)'
Translation Term (TR)

Use the translation term operator (TR) to expand a query to include all defined foreign language equivalent terms.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR(term[, lang, [thes]])</td>
<td>Expands term to include all the foreign equivalents that are defined for term.</td>
</tr>
</tbody>
</table>

**term**

Specify the operand for the translation term operator. term is expanded to include all the foreign language entries defined for term in thes. You cannot specify expansion operators in the term argument.

**lang**

Optionally, specify which foreign language equivalents to return in the expansion. The language you specify must match the language as defined in thes. (You may specify only one language at a time.) If you omit this parameter or specify it as ALL, the system expands to use all defined foreign language terms.

**thes**

Optionally, specify the name of the thesaurus used to return the expansions for the specified term. The thes argument has a default value of DEFAULT. As a result, a thesaurus named DEFAULT must exist in the thesaurus tables before you can use any of the thesaurus operators.

**Note:** If you specify thes, then you must also specify lang.

Examples

Consider a thesaurus MY_THES with the following entries for cat:

cat
  SPANISH: gato
  FRENCH: chat

To search for all documents that contain cat and the spanish translation of cat, enter the following query:

'\tr(cat, spanish, my_thes)'

This query expands to:

'{cat}||{gato}'

Related Topics

Browse a thesaurus using procedures in the CTX_THES package.

**See Also:** CTX_THES.TR in Chapter 13, "CTX_THES Package" for more information on browsing the related terms in your thesaurus
Translation Term Synonym (TRSYN)

Use the translation term operator (TR) to expand a query to include all the defined foreign equivalents of the query term, the synonyms of query term, and the foreign equivalents of the synonyms.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRSYN(term[, lang[, thes]])</td>
<td>Expands term to include foreign equivalents of term, the synonyms of term, and the foreign equivalents of the synonyms.</td>
</tr>
</tbody>
</table>

**term**
Specify the operand for this operator. term is expanded to include all the foreign language entries and synonyms defined for term in thes. You cannot specify expansion operators in the term argument.

**lang**
Optionally, specify which foreign language equivalents to return in the expansion. The language you specify must match the language as defined in thes. If you omit this parameter, the system expands to use all defined foreign language terms.

**thes**
Optionally, specify the name of the thesaurus used to return the expansions for the specified term. The thes argument has a default value of DEFAULT. As a result, a thesaurus named DEFAULT must exist in the thesaurus tables before you can use any of the thesaurus operators.

**Note:** If you specify thes, then you must also specify lang.

Examples

Consider a thesaurus MY_THES with the following entries for cat:

```plaintext
cat
  SPANISH: gato
  FRENCH:  chat
  SYN lion
    SPANISH: leon
```

To search for all documents that contain cat, the spanish equivalent of cat, the synonym of cat, and the spanish equivalent of lion, enter the following query:

`'trsyn(cat, spanish, my_thes)'`

This query expands to:

`'(cat)|(gato)|(lion)|(leon)'`

Related Topics

Browse a thesaurus using procedures in the CTX_THES package.
See Also:  CTX.THES.TRSYN in Chapter 13, "CTX.THES Package" for more information on browsing the translation and synonym terms in your thesaurus
Top Term (TT)

Use the top term operator (TT) to replace a term in a query with the top term that has been defined for the term in the standard hierarchy (Broader Term [BT], Narrower Term [NT]) in a thesaurus. A top term is the broadest conceptual term related to a given query term. For example, a thesaurus might define the following hierarchy:

```
DOG
  BT1 CANINE
  BT2 MAMMAL
  BT3 VERTEBRATE
  BT4 ANIMAL
```

The top term for dog in this thesaurus is animal.

Top terms in the generic (BTG, NTG), partitive (BTP, NTP), and instance (BTI, NTI) hierarchies are not returned.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT(term[thes])</td>
<td>Replaces the specified word in a query with the top term in the standard hierarchy (BT, NT) for term.</td>
</tr>
</tbody>
</table>

**term**

Specify the operand for the top term operator. term is replaced by the top term defined for the term in the specified thesaurus. However, if no TT entries are defined for term, term is not replaced in the query expression and term is the result of the expansion.

You cannot specify expansion operators in the term argument.

**thes**

Specify the name of the thesaurus used to return the expansions for the specified term. The thes argument is optional and has a default value of DEFAULT. A thesaurus named DEFAULT must exist in the thesaurus tables if you use this default value.

Example

The term dog has a top term of animal in the standard hierarchy of a thesaurus. A TT query for dog returns all documents that contain the phrase animal. Documents that contain the word dog are not returned.

Related Topics

Browse your thesaurus using procedures in the CTX_THES package.

See Also:  CTX_THES.TT on page 13-46 for more information on browsing the top terms in your thesaurus
weight (*)

The weight operator multiplies the score by the given factor, topping out at 100 when the score exceeds 100. For example, the query `cat, dog*2` sums the score of `cat` with twice the score of `dog`, topping out at 100 when the score is greater than 100.

In expressions that contain more than one query term, use the weight operator to adjust the relative scoring of the query terms. Reduce the score of a query term by using the weight operator with a number less than 1; increase the score of a query term by using the weight operator with a number greater than 1 and less than 10.

The weight operator is useful in ACCUMulate (,), AND (&), or OR () queries when the expression has more than one query term. With no weighting on individual terms, the score cannot tell which of the query terms occurs the most. With term weighting, you can alter the scores of individual terms and hence make the overall document ranking reflect the terms you are interested in.

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>term*n</code></td>
<td>Returns documents that contain <code>term</code>. Calculates score by multiplying the raw score of <code>term</code> by <code>n</code>, where <code>n</code> is a number from 0.1 to 10.</td>
</tr>
</tbody>
</table>

Examples

Suppose you have a collection of sports articles. You are interested in the articles about Brazilian soccer. It turns out that a regular query on `soccer or Brazil` returns many high ranking articles on US soccer. To raise the ranking of the articles on Brazilian soccer, enter the following query:

'`soccer or Brazil*3`'

Table 3–3 illustrates how the weight operator can change the ranking of three hypothetical documents A, B, and C, which all contain information about soccer. The columns in the table show the total score of four different query expressions on the three documents.

<table>
<thead>
<tr>
<th>Table 3–3 Score Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>soccer</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
</tbody>
</table>

The score in the third column containing the query `soccer or Brazil` is the score of the highest scoring term. The score in the fourth column containing the query `soccer or Brazil*3` is the larger of the score of the first column `soccer` and of the score `Brazil` multiplied by three, `Brazil*3`.

With the initial query of `soccer or Brazil`, the documents are ranked in the order C B A. With the query of `soccer or Brazil*3`, the documents are ranked B C A, which is the preferred ranking.
Weights can be added to multiple terms. The query \( \text{Brazil OR (soccer AND Brazil)}^3 \) will increase the relative scores for documents that contain both soccer and Brazil.
Wildcard characters can be used in query expressions to expand word searches into pattern searches. The wildcard characters are:

<table>
<thead>
<tr>
<th>Wildcard Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>The percent wildcard can appear any number of times at any part of the search term. The search term will be expanded into an equivalence list of terms. The list consists of all terms in the index that match the wildcarded term, with zero or more characters in place of the percent character.</td>
</tr>
<tr>
<td>_</td>
<td>The underscore wildcard specifies a single position in which any character can occur.</td>
</tr>
</tbody>
</table>

The total number of wildcard expansions from all words in a query containing unescaped wildcard characters cannot exceed the maximum number of expansions specified by the BASIC_WORDLIST attribute WILDCARD_MAXTERMS. For more information, see "BASIC_WORDLIST" on page 3-2.

Note: When a wildcard expression translates to a stopword, the stopword is not included in the query and not highlighted by CTX_DOC.HIGHLIGHT or CTX_DOC.MARKUP.

Right-Truncated Queries

Right truncation involves placing the wildcard on the right-hand-side of the search string.

For example, the following query expression finds all terms beginning with the pattern scal:

'scal%'

Left- and Double-Truncated Queries

Left truncation involves placing the wildcard on the left-hand-side of the search string. To find words such as king, wing or sing, write the query as follows:

'__ing'

For all words that end with ing, enter:

'_%ing'

Combine left-truncated and right-truncated searches to create double-truncated searches. The following query finds all documents that contain words that contain the substring %benz%

'_%benz%'

Improving Wildcard Query Performance

Improve wildcard query performance by adding a substring or prefix index.

When your wildcard queries are left- and double-truncated, you can improve query performance by creating a substring index. Substring indexes improve query
performance for all types of left-truncated wildcard searches such as \%ed, \_ing, or \%benz\%.

When your wildcard queries are right-truncated, you can improve performance by creating a prefix index. A prefix index improves query performance for wildcard searches such as \%o\%.

See Also: "BASIC_WORDLIST" on page 2-68 in Chapter 2, "Oracle Text Indexing Elements" for more information about creating substring and prefix indexes.
WITHIN

Use the WITHIN operator to narrow a query down into document sections. Document sections can be one of the following:

- Zone sections
- Field sections
- Attribute sections
- Special sections (sentence or paragraph)

Syntax

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>expression WITHIN section</td>
<td>Searches for expression within the pre-defined zone, field, or attribute section.</td>
</tr>
<tr>
<td></td>
<td>If section is a zone, expression can contain one or more WITHIN operators (nested WITHIN) whose section is a zone or special section.</td>
</tr>
<tr>
<td></td>
<td>If section is a field or attribute section, expression cannot contain another WITHIN operator.</td>
</tr>
<tr>
<td>expression WITHIN SENTENCE</td>
<td>Searches for documents that contain expression within a sentence. Specify an AND or NOT query for expression.</td>
</tr>
<tr>
<td></td>
<td>The expression can contain one or more WITHIN operators (nested WITHIN) whose section is a zone or special section.</td>
</tr>
<tr>
<td>expression WITHIN PARAGRAPH</td>
<td>Searches for documents that contain expression within a paragraph. Specify an AND or NOT query for expression.</td>
</tr>
<tr>
<td></td>
<td>The expression can contain one or more WITHIN operators (nested WITHIN) whose section is a zone or special section.</td>
</tr>
</tbody>
</table>

WITHIN Limitations

The WITHIN operator has the following limitations:

- You cannot embed the WITHIN clause in a phrase. For example, you cannot write: term1 WITHIN section term2
- Because WITHIN is a reserved word, you must escape the word with braces to search on it.

WITHIN Operator Examples

Querying Within Zone Sections
To find all the documents that contain the term San Francisco within the section Headings, write the query as follows:

'San Francisco WITHIN Headings'

To find all the documents that contain the term sailing and contain the term San Francisco within the section Headings, write the query in one of two ways:
Compound Expressions with WITHIN
To find all documents that contain the terms *dog* and *cat* within the same section Headings, write the query as follows:

'(dog and cat) WITHIN Headings'

This query is logically different from:

'dog WITHIN Headings and cat WITHIN Headings'

This query finds all documents that contain *dog* and *cat* where the terms *dog* and *cat* are in Headings sections, regardless of whether they occur in the same Headings section or different sections.

Near with WITHIN
To find all documents in which *dog* is near *cat* within the section Headings, write the query as follows:

'dog near cat WITHIN Headings'

**Note:** The near operator has higher precedence than the WITHIN operator so braces are not necessary in this example. This query is equivalent to (dog near cat) WITHIN Headings.

Nested WITHIN Queries
You can nest the within operator to search zone sections within zone sections.

For example, assume that a document set had the zone section AUTHOR nested within the zone BOOK section. Write a nested WITHIN query to find all occurrences of *scott* within the AUTHOR section of the BOOK section as follows:

'(scott WITHIN AUTHOR) WITHIN BOOK'

Querying Within Field Sections
The syntax for querying within a field section is the same as querying within a zone section. The syntax for most of the examples given in the previous section, "Querying Within Zone Sections", apply to field sections.

However, field sections behave differently from zone sections in terms of

- Visibility: Make text within a field section invisible.
- Repeatability: WITHIN queries cannot distinguish repeated field sections.
- Nestability: You cannot enter a nested WITHIN query with a field section.

The following sections describe these differences.

Visible Flag in Field Sections
When a field section is created with the visible flag set to FALSE in CTX_DDL.ADD_FIELD_SECTION, the text within a field section can only be queried using the WITHIN operator.
For example, assume that TITLE is a field section defined with visible flag set to FALSE. Then the query dog without the WITHIN operator will not find a document containing:

<TITLE>The dog</TITLE> I like my pet.

To find such a document, use the WITHIN operator as follows:

'dog WITHIN TITLE'

Alternatively, set the visible flag to TRUE when you define TITLE as a field section with CTX_DDL.ADD_FIELD_SECTION.

See Also: "ADD_FIELD_SECTION" in Chapter 7, "CTX_DDL Package" for more information about creating field sections

Repeated Field Sections

WITHIN queries cannot distinguish repeated field sections in a document. For example, consider the document with the repeated section <author>:

<author> Charles Dickens </author>
<author> Martin Luther King </author>

Assuming that <author> is defined as a field section, a query such as (charles and martin) within author returns the document, even though these words occur in separate tags.

To have WITHIN queries distinguish repeated sections, define the sections as zone sections.

Nested Field Sections

You cannot enter a nested WITHIN query with field sections. Doing so raises an error.

Querying Within Sentence or Paragraphs

Querying within sentence or paragraph boundaries is useful to find combinations of words that occur in the same sentence or paragraph. To query sentence or paragraphs, you must first add the special section to your section group before you index. Do so with CTX_DDL.ADD_SPECIAL_SECTION.

To find documents that contain dog and cat within the same sentence:

'(dog and cat) WITHIN SENTENCE'

To find documents that contain dog and cat within the same paragraph:

'(dog and cat) WITHIN PARAGRAPH'

To find documents that contain sentences with the word dog but not cat:

'(dog not cat) WITHIN SENTENCE'

Querying Within Attribute Sections

Query within attribute sections when you index with either XML_SECTION_GROUP or AUTO_SECTION_GROUP as your section group type.

Assume you have an XML document as follows:

<book title='Tale of Two Cities'>It was the best of times.</book>
Define the section title@book to be the attribute section title. Do so with the CTX_DLL.ADD_ATTR_SECTION procedure or dynamically after indexing with ALTER INDEX.

---

**Note:** When you use the AUTO_SECTION_GROUP to index XML documents, the system automatically creates attribute sections and names them in the form attribute@tag.

If you use the XML_SECTION_GROUP, you can name attribute sections anything with CTX_DDL.ADD_ATTR_SECTION.

---

To search on Tale within the attribute section title, enter the following query:

'Tale WITHIN title'

**Constraints for Querying Attribute Sections**

The following constraints apply to querying within attribute sections:

- Regular queries on attribute text do not hit the document unless qualified in a within clause. Assume you have an XML document as follows:

```xml
<book title="Tale of Two Cities">It was the best of times.</book>
```

A query on Tale by itself does not produce a hit on the document unless qualified with WITHIN title@book. (This behavior is like field sections when you set the visible flag set to false.)

- You cannot use attribute sections in a nested WITHIN query.

- Phrases ignore attribute text. For example, if the original document looked like:

```xml
<book title="Tale of Two Cities">It was the best of times.</book>
```

Now is the time for all good <word type="noun">men</word> to come to the aid.

Then this document would hit on the regular query good men, ignoring the intervening attribute text.

- WITHIN queries can distinguish repeated attribute sections. This behavior is like zone sections but unlike field sections. For example, you have a document as follows:

```xml
<book title="Tale of Two Cities">It was the best of times.</book>
<book title="Of Human Bondage">The sky broke dull and gray.</book>
```

Assume that book is a zone section and book@author is an attribute section. Consider the query:

'\(\text{Tale and Bondage}) \text{ WITHIN book@author}''

This query does *not* hit the document, because *tale* and *bondage* are in different occurrences of the attribute section book@author.

---

**Notes**

**Section Names**

The WITHIN operator requires you to know the name of the section you search. A list of defined sections can be obtained using the CTX_SECTIONS or CTX_USER_SECTIONS views.
Section Boundaries

For special and zone sections, the terms of the query must be fully enclosed in a particular occurrence of the section for the document to satisfy the query. This is not a requirement for field sections.

For example, consider the query where bold is a zone section:

'(dog and cat) WITHIN bold'

This query finds:

<B>dog cat</B>

but it does not find:

<B>dog</B><B>cat</B>

This is because dog and cat must be in the same bold section.

This behavior is especially useful for special sections, where

'(dog and cat) WITHIN sentence'

means find dog and cat within the same sentence.

Field sections on the other hand are meant for non-repeating, embedded metadata such as a title section. Queries within field sections cannot distinguish between occurrences. All occurrences of a field section are considered to be parts of a single section. For example, the query:

(dog and cat) WITHIN title

can find a document like this:

<TITLE>dog</TITLE><TITLE>cat</TITLE>

In return for this field section limitation and for the overlap and nesting limitations, field section queries are generally faster than zone section queries, especially if the section occurs in every document, or if the search term is common.
This chapter describes the special characters that can be used in Text queries. In addition, it provides a list of the words and characters that Oracle Text treats as reserved words and characters.

The following topics are covered in this chapter:

- Grouping Characters
- Escape Characters
- Reserved Words and Characters

**Grouping Characters**

The grouping characters control operator precedence by grouping query terms and operators in a query expression. The grouping characters are:

<table>
<thead>
<tr>
<th>Grouping Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>()</td>
<td>The parentheses characters serve to group terms and operators found between the characters</td>
</tr>
<tr>
<td>[]</td>
<td>The bracket characters serve to group terms and operators found between the characters; however, they prevent penetrations for the expansion operators (fuzzy, soundex, stem).</td>
</tr>
</tbody>
</table>

The beginning of a group of terms and operators is indicated by an open character from one of the sets of grouping characters. The ending of a group is indicated by the occurrence of the appropriate close character for the open character that started the group. Between the two characters, other groups may occur.

For example, the open parenthesis indicates the beginning of a group. The first close parenthesis encountered is the end of the group. Any open parentheses encountered before the close parenthesis indicate nested groups.

**Escape Characters**

To query on words or symbols that have special meaning to query expressions such as *and* & or | accum, you must escape them. There are two ways to escape characters in a query expression:
In the following examples, an escape sequence is necessary because each expression contains a Text operator or reserved symbol:

'high\-voltage'

'{high-voltage}'

'XY\&Z'

'{XY&Z}'

In the first example, the query matches *high-voltage* or *high voltage*.

Note that in the second example, a query on *XY&Z* will return 'XY Z', 'XY-Z', 'XY*Z', and so forth, as well as 'XY&Z'. This is because non-alphabetic characters are treated as whitespace (so *XY&Z* is treated as 'XY Z'). To match only *XY&Z*, you must declare & as a printjoin. (If you do, however, *XY&Z* will not match 'XY & Z'.) For more on printjoins, see BASIC_LEXER on page 2-39.

<table>
<thead>
<tr>
<th>Escape Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>{}</td>
<td>Use braces to escape a string of characters or symbols. Everything within a set of braces is considered part of the escape sequence. When you use braces to escape a single character, the escaped character becomes a separate token in the query.</td>
</tr>
<tr>
<td>\</td>
<td>Use the backslash character to escape a single character or symbol. Only the character immediately following the backslash is escaped. For example, a query of <em>blue-green</em> matches <em>blue-green</em> and <em>blue green</em>.</td>
</tr>
</tbody>
</table>

Note: If you use braces to escape an individual character within a word, the character is escaped, but the word is broken into three tokens.

For example, a query written as *high\-voltage* searches for *high-voltage*, with the space on either side of the hyphen.

**Querying Escape Characters**

The open brace { signals the beginning of the escape sequence, and the closed brace } indicates the end of the sequence. Everything between the opening brace and the closing brace is part of the escaped query expression (including any open brace characters). To include the close brace character in an escaped query expression, use } }.

To escape the backslash escape character, use \\.

**Reserved Words and Characters**

Table 4–3 lists the Oracle Text reserved words and characters that must be escaped when you want to search them in CONTAINS queries:

<table>
<thead>
<tr>
<th>Reserved Words</th>
<th>Reserved Characters</th>
<th>Operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABOUT</td>
<td>(none)</td>
<td>ABOUT</td>
</tr>
</tbody>
</table>
### Table 4-3  (Cont.) Reserved Words and Characters

<table>
<thead>
<tr>
<th>Reserved Words</th>
<th>Reserved Characters</th>
<th>Operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCUM</td>
<td>,</td>
<td>Accumulate</td>
</tr>
<tr>
<td>AND</td>
<td>&amp;</td>
<td>And</td>
</tr>
<tr>
<td>BT</td>
<td>(none)</td>
<td>Broader Term</td>
</tr>
<tr>
<td>BTG</td>
<td>(none)</td>
<td>Broader Term Generic</td>
</tr>
<tr>
<td>BTI</td>
<td>(none)</td>
<td>Broader Term Instance</td>
</tr>
<tr>
<td>BTP</td>
<td>(none)</td>
<td>Broader Term Partitive</td>
</tr>
<tr>
<td>EQUIV</td>
<td>=</td>
<td>Equivalence</td>
</tr>
<tr>
<td>FUZZY</td>
<td>?</td>
<td>fuzzy</td>
</tr>
<tr>
<td>(none)</td>
<td>{ }</td>
<td>escape characters (multiple)</td>
</tr>
<tr>
<td>(none)</td>
<td>\</td>
<td>escape character (single)</td>
</tr>
<tr>
<td>(none)</td>
<td>( )</td>
<td>grouping characters</td>
</tr>
<tr>
<td>(none)</td>
<td>[ ]</td>
<td>grouping characters</td>
</tr>
<tr>
<td>HASPATH</td>
<td>(none)</td>
<td>HASPATH</td>
</tr>
<tr>
<td>INPATH</td>
<td>(none)</td>
<td>INPATH</td>
</tr>
<tr>
<td>MDATA</td>
<td>(none)</td>
<td>MDATA</td>
</tr>
<tr>
<td>MINUS</td>
<td>-</td>
<td>MINUS</td>
</tr>
<tr>
<td>NEAR</td>
<td>;</td>
<td>NEAR</td>
</tr>
<tr>
<td>NOT</td>
<td>~</td>
<td>NOT</td>
</tr>
<tr>
<td>NT</td>
<td>(none)</td>
<td>Narrower Term</td>
</tr>
<tr>
<td>NTG</td>
<td>(none)</td>
<td>Narrower Term Generic</td>
</tr>
<tr>
<td>NTI</td>
<td>(none)</td>
<td>Narrower Term Instance</td>
</tr>
<tr>
<td>NTP</td>
<td>(none)</td>
<td>Narrower Term Partitive</td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>(none)</td>
<td>Preferred Term</td>
</tr>
<tr>
<td>RT</td>
<td>(none)</td>
<td>Related Term</td>
</tr>
<tr>
<td>(none)</td>
<td>$</td>
<td>stem</td>
</tr>
<tr>
<td>(none)</td>
<td>!</td>
<td>soundex</td>
</tr>
<tr>
<td>SQE</td>
<td>(none)</td>
<td>Stored Query Expression</td>
</tr>
<tr>
<td>SYN</td>
<td>(none)</td>
<td>Synonym</td>
</tr>
<tr>
<td>(none)</td>
<td>&gt;</td>
<td>threshold</td>
</tr>
<tr>
<td>TR</td>
<td>(none)</td>
<td>Translation Term</td>
</tr>
<tr>
<td>TRSYN</td>
<td>(none)</td>
<td>Translation Term Synonym</td>
</tr>
<tr>
<td>TT</td>
<td>(none)</td>
<td>Top Term</td>
</tr>
<tr>
<td>(none)</td>
<td>*</td>
<td>weight</td>
</tr>
<tr>
<td>(none)</td>
<td>%</td>
<td>wildcard character (multiple)</td>
</tr>
<tr>
<td>(none)</td>
<td>_</td>
<td>wildcard character (single)</td>
</tr>
<tr>
<td>Reserved Words</td>
<td>Reserved Characters</td>
<td>Operator</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------</td>
<td>----------</td>
</tr>
<tr>
<td>WITHIN</td>
<td>(none)</td>
<td>WITHIN</td>
</tr>
</tbody>
</table>
This chapter provides information for using the CTX_ADM PL/SQL package.

CTX_ADM contains the following stored procedures:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARK_FAILED</td>
<td>Changes an index’s status from LOADING to FAILED.</td>
</tr>
<tr>
<td>RECOVER</td>
<td>Cleans up database objects for deleted Text tables.</td>
</tr>
<tr>
<td>SET_PARAMETER</td>
<td>Sets system-level defaults for index creation.</td>
</tr>
</tbody>
</table>

**Note:** Only the CTXSYS user can use the procedures in CTX_ADM.
MARK_FAILED

Use this procedure to change the status of an index from LOADING to FAILED.

Under rare circumstances, if CREATE INDEX or ALTER INDEX fails, an index may be left with the status LOADING. When an index is in LOADING status, any attempt to recover using RESUME INDEX is blocked. For this situation, use CTX_ADM.MARK_FAILED to forcibly change the status from LOADING to FAILED so that you can recover the index with RESUME INDEX.

You must log on as CTXSYS to run CTX_ADM.MARK_FAILED.

CAUTION: Use CTX_ADM.MARK_FAILED with caution. It should only be used as a last resort and only when no other session is touching the index. Normally, CTX_ADM.MARK_FAILED does not succeed if another session is actively building the index with CREATE or ALTER INDEX. However, index creation or alteration may include windows of time during which CTX_ADM.MARK_FAILED can succeed, marking the index as failed even as it is being built by another session.

CTX_ADM.MARK_FAILED works with local partitioned indexes. However, it changes the status of all partitions to FAILED. Therefore, you should rebuild all index partitions with ALTER INDEX REBUILD PARTITION PARAMETERS ('RESUME') after using CTX_ADM.MARK_FAILED. If you run ALTER INDEX PARAMETER ('RESUME') after this operation, then Oracle resets the index partition status to valid. Oracle does not rebuild the index partitions that were successfully built before the MARK_FAILED operation.

Syntax

CTX_ADM.MARK_FAILED(
    owner_name     in    VARCHAR2,
    index_name     in    VARCHAR2);

owner_name
The name of the owner of the index whose status is to be changed.

index_name
The name of the index whose status is to be changed.

Example

begin
    CTX_ADM.MARK_FAILED('owner_1', 'index_1');
end;
RECOVER

The RECOVER procedure cleans up the Text data dictionary, deleting objects such as leftover preferences.

Syntax

CTX_ADM.RECOVER;

Example

begin
  ctx_adm.recover;
end;
The **SET_PARAMETER** procedure sets system-level parameters for index creation.

**Syntax**

```sql
CTX_ADM.SET_PARAMETER(param_name IN VARCHAR2,
                         param_value IN VARCHAR2);
```

**param_name**
Specify the name of the parameter to set, which can be one of the following parameters:

- `max_index_memory` (maximum memory allowed for indexing)
- `default_index_memory` (default memory allocated for indexing)
- `log_directory` (directory for CTX_OUTPUT files)
- `ctx_doc_key_type` (default input key type for CTX_DOC procedures)
- `file_access_role` (default database role name for index creation when using FILE or URL datastores)
- `default_datastore` (default datastore preference)
- `default_filter_file` (default filter preference for data stored in files)
- `default_filter_text` (default text filter preference)
- `default_filter_binary` (default binary filter preference)
- `default_section_html` (default html section group preference)
- `default_section_xml` (default xml section group preference)
- `default_section_text` (default text section group preference)
- `default_lexer` (default lexer preference)
- `default_wordlist` (default wordlist preference)
- `default_stoplist` (default stoplist preference)
- `default_storage` (default storage preference)
- `default_ctxcat_lexer`
- `default_ctxcat_stoplist`
- `default_ctxcat_storage`
- `default_ctxcat_wordlist`
- `default_ctxrule_lexer`
- `default_ctxrule_stoplist`
- `default_ctxrule_storage`
- `default_ctxrule_wordlist`

**See Also:** To learn more about the default values for these parameters, see "System Parameters" on page 2-87 in Chapter 2, "Oracle Text Indexing Elements"
**param_value**
Specify the value to assign to the parameter. For `max_index_memory` and `default_index_memory`, the value you specify must have the following syntax:

```
number[K|M|G]
```

where K stands for kilobytes, M stands for megabytes, and G stands for gigabytes.
For each of the other parameters, specify the name of a preference to use as the default for indexing.

**Example**

```python
begin
    ctx_adm.set_parameter('default_lexer', 'my_lexer');
end;
```
SET_PARAMETER
This chapter provides reference information for using the CTX_CLS PL/SQL package, which enables you to perform document classification.

The following procedures are described in this chapter:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLUSTERING</td>
<td>Generates clusters for a document collection.</td>
</tr>
</tbody>
</table>

**See Also:** Oracle Text Application Developer’s Guide for more on document classification
Use this procedure to generate query rules that select document categories. You must supply a training set consisting of categorized documents. Documents can be in any format supported by Oracle Text and must belong to one or more categories. This procedure generates the queries that define the categories and then writes the results to a table.

You must also have a document table and a category table. The category table must contain at least two categories.

For example, your document and category tables can be defined as:

```sql
create table trainingdoc(
    docid number primary key,
    text varchar2(4000)));
create table category {
    docid trainingdoc(docid),
    categoryid number);
```

You can use one of two syntaxes depending on the classification algorithm you need. The query compatible syntax uses the RULE_CLASSIFIER preference and generates rules as query strings. The support vector machine syntax uses the SVM_CLASSIFIER preference and generates rules in binary format. The SVM_CLASSIFIER is good for high classification accuracy, but because its rules are generated in binary format, they cannot be examined like the query strings generated with the RULE_CLASSIFIER. Note that only those document ids that appear in both the document table and the category table will impact RULE_CLASSIFIER and SVM_CLASSIFIER learning.

The CTX_CLS.TRAIN procedure requires that your document table have an associated context index. For best results, the index should be synchronized before running this procedure. SVM_CLASSIFIER syntax enables the use of an unpopulated context index, while query-compatible syntax requires that the context index be populated.

**See Also:** Oracle Text Application Developer’s Guide for more on document classification

### Query Compatible Syntax

The following syntax generates query-compatible rules and is used with the RULE_CLASSIFIER preference. Use this syntax and preference when different categories are separated from others by several key words. An advantage of generating your rules as query strings is that you can easily examine the generated rules. This is different from generating SVM rules, which are in binary format.

```sql
CTX_CLS.TRAIN(
    index_name    in varchar2,
    docid         in varchar2,
    cattab        in varchar2,
    catdocid      in varchar2,
    catid         in varchar2,
    restab        in varchar2,
    rescatid      in varchar2,
    resquery      in varchar2,
    resconfid     in varchar2,
    preference    in varchar2 DEFAULT NULL
);
```
index_name
Specify the name of the context index associated with your document training set.

docid
Specify the name of the document ID column in the document table. The document IDs in this column must be unique, and this column must be of datatype NUMBER. The values for this column must be stored in an unsigned 32-bit integer and must be in the range 0-4294967295.

cattab
Specify the name of the category table. You must have SELECT privilege on this table.

catdocid
Specify the name of the document ID column in the category table. The document IDs in this table must also exist in the document table. This column must be NUMBER. The values for this column must be stored in an unsigned 32-bit integer and must be in the range 0-4294967295.

catid
Specify the name of the category ID column in the category table. This column must be NUMBER. The values for this column must be stored in an unsigned 32-bit integer and must be in the range 0-4294967295.

restab
Specify the name of the result table. You must have INSERT privilege on this table.

rescatid
Specify the name of the category ID column in the result table. This column must be NUMBER. The values for this column must be stored in an unsigned 32-bit integer and must be in the range 0-4294967295.

resquery
Specify the name of the query column in the result table. This column must be VARCHAR2, CHAR, CLOB, NVARCHAR2, or NCHAR.
The queries generated in this column connects terms with AND or NOT operators, such as:

'T1 & T2 ~ T3'

Terms can also be theme tokens and be connected with the ABOUT operator, such as:

'about(T1) & about(T2) ~ about(T3)'

Generated rules also support WITHIN queries on field sections.

resconfid
Specify the name of the confidence column in result table. This column contains the estimated probability from training data that a document is relevant if that document satisfies the query.

preference
Specify the name of the preference. For classifier types and attributes, see "Classifier Types" on page 2-79 in Chapter 2, "Oracle Text Indexing Elements".
Syntax for Support Vector Machine Rules

The following syntax generates support vector machine (SVM) rules with the SVM_CLASSIFIER preference. This preference generates rules in binary format. Use this syntax when your application requires high classification accuracy.

```sql
CTX_CLS.TRAIN(
    index_name in varchar2,
    docid in varchar2,
    cattab in varchar2,
    catdocid in varchar2,
   catid in varchar2,
    restab in varchar2,
    preference in varchar2 );
```

**index_name**
Specify the name of the text index.

**docid**
Specify the name of docid column in document table.

**cattab**
Specify the name of category table.

**catdocid**
Specify the name of docid column in category table.

**catid**
Specify the name of category ID column in category table.

**restab**
Specify the name of result table.

The result table has the following format:

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Datatype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT_ID</td>
<td>NUMBER</td>
<td>The ID of the category.</td>
</tr>
<tr>
<td>TYPE</td>
<td>NUMBER(3) NOT NULL</td>
<td>0 for the actual rule or catid; 1 for other.</td>
</tr>
<tr>
<td>RULE</td>
<td>BLOB</td>
<td>The returned rule.</td>
</tr>
</tbody>
</table>

**preference**
Specify the name of user preference. For classifier types and attributes, see "Classifier Types" on page 2-79 in Chapter 2, "Oracle Text Indexing Elements".

**Example**

The CTX_CLS.TRAIN procedure is used in supervised classification. For an extended example, see Oracle Text Application Developer’s Guide.
CLUSTERING

Use this procedure to cluster a collection of documents. A cluster is a group of
documents similar to each other in content.

A clustering result set is composed of document assignments and cluster descriptions:

- A document assignment result set shows how relevant each document is to all
generated leaf clusters.
- A cluster description result set contains information about what topic a cluster is
about. This result set identifies the cluster and contains cluster description text, a
suggested cluster label, and a quality score for the cluster.

Cluster output is hierarchical. Only leaf clusters are scored for relevance to documents.
Producing more clusters requires more computing time. Indicate the upper limit for
generated clusters with the CLUSTER_NUM attribute of the KMEAN_CLUSTERING
cluster type (see "Cluster Types" on page 2-82 in this chapter).

There are two versions of this procedure: one with a table result set, and one with an
in-memory result set.

Clustering is also known as unsupervised classification.

See Also: For more information about clustering and relevant
preferences, see Cluster Types on page 2-82 in Chapter 2, "Oracle
Text Indexing Elements", as well as the Oracle Text Application
Developer’s Guide

Syntax: Table Result Set

cctx_cls.clustering (  
  index_name  IN VARCHAR2,  
  docid       IN VARCHAR2,  
  doctab_name IN VARCHAR2,  
  clstab_name IN VARCHAR2,  
  pref_name   IN VARCHAR2  DEFAULT NULL  
);

index_name
Specify the name of the context index on collection table.

docid
Specify the name of document ID column of the collection table.

doctab_name
Specify the name of document assignment table. This procedure creates the table with
the following structure:

doc_assign(  
  docid number,  
  clusterid number,  
  score number  
);

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOCID</td>
<td>Document ID to identify document.</td>
</tr>
</tbody>
</table>
If you require more columns, then create the table before you call this procedure.

**clstab_name**
Specify the name of the cluster description table. This procedure creates the table with the following structure:

```sql
cluster_desc(
    clusterid NUMBER,
    descript VARCHAR2(4000),
    label VARCHAR2(200),
    sze NUMBER,
    quality_score NUMBER,
    parent NUMBER
);
```

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLUSTERID</td>
<td>Cluster ID to identify cluster. If CLUSTERID is -1, then the cluster contains &quot;miscellaneous&quot; documents; for example, documents that cannot be assigned to any other cluster category.</td>
</tr>
<tr>
<td>DESCRIPT</td>
<td>String to describe the cluster.</td>
</tr>
<tr>
<td>LABEL</td>
<td>A suggested label for the cluster.</td>
</tr>
<tr>
<td>SZE</td>
<td>This parameter currently has no value.</td>
</tr>
<tr>
<td>QUALITY_SCORE</td>
<td>The quality score of the cluster. A higher number indicates greater coherence.</td>
</tr>
<tr>
<td>PARENT</td>
<td>The parent cluster ID. Zero means no parent cluster.</td>
</tr>
</tbody>
</table>

If you require more columns, then create the table before you call this procedure.

**pref_name**
Specify the name of the preference.

**Syntax: In-Memory Result Set**

Put the result set into in-memory structures for better performance. Two in-memory tables are defined in CTX_CLS package for document assignment and cluster description respectively.

```sql
CTX_CLS.CLUSTERING(
    index_name     IN VARCHAR2,
    docid          IN VARCHAR2,
    dids           IN DOCID_TAB,
    doctab_name    IN OUT NOCOPY DOC_TAB,
    clstab_name    IN OUT NOCOPY CLUSTER_TAB,
    pref_name      IN VARCHAR2  DEFAULT NULL
);
```
index_name
Specify the name of context index on the collection table.

docid
Specify the document ID column of the collection table.

dids
Specify the name of the in-memory docid_tab.

TYPE docid_tab IS TABLE OF number INDEX BY BINARY_INTEGER;

doctab_name
Specify name of the document assignment in-memory table. This table is defined as follows:

TYPE doc_rec IS RECORD (
   docid NUMBER,
   clusterid NUMBER,
   score NUMBER
);
TYPE doc_tab IS TABLE OF doc_rec INDEX BY BINARY_INTEGER;

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOCID</td>
<td>Document ID to identify document.</td>
</tr>
<tr>
<td>CLUSTERID</td>
<td>ID of a leaf cluster associated with this document. If CLUSTERID is -1, then the cluster contains &quot;miscellaneous&quot; documents; for example, documents that cannot be assigned to any other cluster category.</td>
</tr>
<tr>
<td>SCORE</td>
<td>The associated score between the document and the cluster.</td>
</tr>
</tbody>
</table>

cls_tab
Specify the name of cluster description in-memory table.

TYPE cluster_rec IS RECORD(
   clusterid NUMBER,
   descrip VARCHAR2(4000),
   label VARCHAR2(200),
   sze NUMBER,
   quality_score NUMBER,
   parent NUMBER
);
TYPE cluster_tab IS TABLE OF cluster_rec INDEX BY BINARY_INTEGER;

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLUSTERID</td>
<td>Cluster ID to identify cluster. If CLUSTERID is -1, then the cluster contains &quot;miscellaneous&quot; documents; for example, documents that cannot be assigned to any other cluster category.</td>
</tr>
<tr>
<td>DESCRIPT</td>
<td>String to describe the cluster.</td>
</tr>
<tr>
<td>LABEL</td>
<td>A suggested label for the cluster.</td>
</tr>
<tr>
<td>SZE</td>
<td>This parameter currently has no value.</td>
</tr>
<tr>
<td>QUALITY_SCORE</td>
<td>The quality score of the cluster. A higher number indicates greater coherence.</td>
</tr>
<tr>
<td>PARENT</td>
<td>The parent cluster ID. Zero means no parent cluster.</td>
</tr>
</tbody>
</table>
**pref_name**
Specify the name of the preference. For cluster types and attributes, see Cluster Types in Chapter 2, "Oracle Text Indexing Elements".

**Example**

**See Also:** The *Oracle Text Application Developer’s Guide* for an example of using clustering
This chapter provides reference information for using the CTX_DDL PL/SQL package to create and manage the preferences, section groups, and stoplists required for Text indexes.

CTX_DDL contains the following stored procedures and functions:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD_ATTR_SECTION</td>
<td>Adds an attribute section to an XML section group.</td>
</tr>
<tr>
<td>ADD_FIELD_SECTION</td>
<td>Creates a field section and assigns it to the specified section group.</td>
</tr>
<tr>
<td>ADD_INDEX</td>
<td>Adds an index to a catalog index preference.</td>
</tr>
<tr>
<td>ADD_MDATA</td>
<td>Changes the MDATA value of a document.</td>
</tr>
<tr>
<td>ADD_MDATA_COLUMN</td>
<td>Maps a FILTER BY column to the specified MDATA section.</td>
</tr>
<tr>
<td>ADD_MDATA_SECTION</td>
<td>Adds an MDATA metadata section to a document.</td>
</tr>
<tr>
<td>ADD_NDATA_SECTION</td>
<td>Adds an NDATA section to a document.</td>
</tr>
<tr>
<td>ADD_SDATA_COLUMN</td>
<td>Maps a FILTER BY column to the specified SDATA section.</td>
</tr>
<tr>
<td>ADD_SDATA_SECTION</td>
<td>Adds an SDATA structured data section to a document.</td>
</tr>
<tr>
<td>ADD_SPECIAL_SECTION</td>
<td>Adds a special section to a section group.</td>
</tr>
<tr>
<td>ADD_STOPCLASS</td>
<td>Adds a stopclass to a stoplist.</td>
</tr>
<tr>
<td>ADD_STOP SECTION</td>
<td>Adds a stop section to an automatic section group.</td>
</tr>
<tr>
<td>ADD_STOPTHEME</td>
<td>Adds a stoptheme to a stoplist.</td>
</tr>
<tr>
<td>ADD_STOPWORD</td>
<td>Adds a stopword to a stoplist.</td>
</tr>
<tr>
<td>ADD_SUB_LEXER</td>
<td>Adds a sub-lexer to a multi-lexer preference.</td>
</tr>
<tr>
<td>ADD_ZONE_SECTION</td>
<td>Creates a zone section and adds it to the specified section group.</td>
</tr>
<tr>
<td>COPY_POLICY</td>
<td>Creates a copy of a policy.</td>
</tr>
<tr>
<td>CREATE_INDEX_SET</td>
<td>Creates an index set for CTXCAT index types.</td>
</tr>
<tr>
<td>CREATE_POLICY</td>
<td>Creates a policy to use with ORA:CONTAINS().</td>
</tr>
<tr>
<td>CREATE_PREFERENCE</td>
<td>Creates a preference in the Text data dictionary.</td>
</tr>
<tr>
<td>CREATE_SECTION_GROUP</td>
<td>Creates a section group in the Text data dictionary.</td>
</tr>
<tr>
<td>Name</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>CREATE_SHADOW_INDEX</td>
<td>Creates a policy for the passed-in index. For non-partitioned index, also creates an index table.</td>
</tr>
<tr>
<td>CREATE_STOPLIST</td>
<td>Creates a stoplist.</td>
</tr>
<tr>
<td>DROP_INDEX_SET</td>
<td>Drops an index set.</td>
</tr>
<tr>
<td>DROP_POLICY</td>
<td>Drops a policy.</td>
</tr>
<tr>
<td>DROP_PREFERENCE</td>
<td>Deletes a preference from the Text data dictionary.</td>
</tr>
<tr>
<td>DROP_SECTION_GROUP</td>
<td>Deletes a section group from the Text data dictionary.</td>
</tr>
<tr>
<td>DROP_SHADOW_INDEX</td>
<td>Drops a shadow index.</td>
</tr>
<tr>
<td>DROP_STOPLIST</td>
<td>Drops a stoplist.</td>
</tr>
<tr>
<td>EXCHANGE_SHADOW_INDEX</td>
<td>Swaps the shadow index metadata and data.</td>
</tr>
<tr>
<td>OPTIMIZE_INDEX</td>
<td>Optimizes the index.</td>
</tr>
<tr>
<td>POPULATE_PENDING</td>
<td>Populates the pending queue with every rowid in the base table or table partition.</td>
</tr>
<tr>
<td>RECREATE_INDEX_ONLINE</td>
<td>Recreates the passed-in index.</td>
</tr>
<tr>
<td>REMOVE_INDEX</td>
<td>Removes an index from a CTXCAT index preference.</td>
</tr>
<tr>
<td>REMOVE_MDATA</td>
<td>Removes MDATA values from a document.</td>
</tr>
<tr>
<td>REMOVE_SECTION</td>
<td>Deletes a section from a section group.</td>
</tr>
<tr>
<td>REMOVE_STOPCLASS</td>
<td>Deletes a stopclass from a stoplist.</td>
</tr>
<tr>
<td>REMOVE_STOPTHEME</td>
<td>Deletes a stoptheme from a stoplist.</td>
</tr>
<tr>
<td>REMOVE_STOPWORD</td>
<td>Deletes a stopword from a stoplist.</td>
</tr>
<tr>
<td>REMOVE_SUB_LEXER</td>
<td>Deletes a sub-lexer from a multi-lexer preference.</td>
</tr>
<tr>
<td>REPLACE_INDEX_METADATA</td>
<td>Replaces metadata for local domain indexes.</td>
</tr>
<tr>
<td>SET_ATTRIBUTE</td>
<td>Sets a preference attribute.</td>
</tr>
<tr>
<td>SYNC_INDEX</td>
<td>Synchronizes the index.</td>
</tr>
<tr>
<td>UNSET_ATTRIBUTE</td>
<td>Removes a set attribute from a preference.</td>
</tr>
<tr>
<td>UPDATE_POLICY</td>
<td>Updates a policy.</td>
</tr>
</tbody>
</table>
**ADD_ATTR_SECTION**

Adds an attribute section to an XML section group. This procedure is useful for defining attributes in XML documents as sections. This enables you to search XML attribute text with the `WITHIN` operator.

**Syntax**

```sql
CTX_DDL.ADD_ATTR_SECTION(
    group_name     in    varchar2,
    section_name   in    varchar2,
    tag            in    varchar2);
```

- **group_name**
  Specify the name of the XML section group. You can add attribute sections only to XML section groups.

- **section_name**
  Specify the name of the attribute section. This is the name used for `WITHIN` queries on the attribute text.
  The section name you specify cannot contain the colon (:), comma (,), or dot (.) characters. The section name must also be unique within `group_name`. Section names are case-insensitive.
  Attribute section names can be no more than 64 bytes long.

- **tag**
  Specify the name of the attribute in `tag@attr` form. This parameter is case-sensitive.

**Examples**

Consider an XML file that defines the `BOOK` tag with a `TITLE` attribute as follows:

```xml
<Book TITLE="Tale of Two Cities">
  It was the best of times.
</Book>
```

To define the title attribute as an attribute section, create an `XML_SECTION_GROUP` and define the attribute section as follows:

```sql
begin
    ctx_ddl.create_section_group('myxmlgroup', 'XML_SECTION_GROUP');
    ctx_ddl.add_attr_section('myxmlgroup', 'booktitle', 'BOOK@TITLE');
end;
```

When you define the `TITLE` attribute section as such and index the document set, you can query the XML attribute text as follows:

`Cities within booktitle`
ADD_FIELD_SECTION

Creates a field section and adds the section to an existing section group. This enables field section searching with the WITHIN operator.

Field sections are delimited by start and end tags. By default, the text within field sections are indexed as a sub-document separate from the rest of the document.

Unlike zone sections, field sections cannot nest or overlap. As such, field sections are best suited for non-repeating, non-overlapping sections such as TITLE and AUTHOR markup in e-mail- or news-type documents.

Because of how field sections are indexed, WITHIN queries on field sections are usually faster than WITHIN queries on zone sections.

Syntax

```
CTX_DDL.ADD_FIELD_SECTION(
    group_name     in    varchar2,
    section_name   in    varchar2,
    tag            in    varchar2,
    visible        in    boolean default FALSE
);
```

**group_name**

Specify the name of the section group to which `section_name` is added. You can add up to 64 field sections to a single section group. Within the same group, section zone names and section field names cannot be the same.

**section_name**

Specify the name of the section to add to the `group_name`. Use this name to identify the section in queries. Avoid using names that contain non-alphanumeric characters such as `_` because these characters must be escaped in queries. Section names are case-insensitive.

Within the same group, zone section names and field section names cannot be the same. The terms Paragraph and Sentence are reserved for special sections.

Section names need not be unique across tags. You can assign the same section name to more than one tag, which makes details transparent to searches.

**tag**

Specify the tag that marks the start of a section. For example, if the tag is `<H1>`, then specify H1. The start tag you specify must be unique within a section group.

If `group_name` is an HTML_SECTION_GROUP, then you can create field sections for the META tag's NAME/CONTENT attribute pairs. To do so, specify tag as meta@namevalue where namevalue is the value of the NAME attribute whose CONTENT attribute is to be indexed as a section. Refer to the example "Creating Sections for `<META>` Tags" on page 7-5.

Oracle Text knows what the end tags look like from the `group_type` parameter you specify when you create the section group.

**visible**

Specify TRUE to make the text visible within the rest of the document.

By default the `visible` flag is FALSE. This means that Oracle Text indexes the text within field sections as a sub-document separate from the rest of the document.
However, you can set the visible flag to TRUE if you want text within the field section to be indexed as part of the enclosing document.

Examples

Visible and Invisible Field Sections
The following example defines a section group basicgroup of the BASIC_SECTION_GROUP type. It then creates a field section in basicgroup called Author for the <A> tag. It also sets the visible flag to FALSE:

```plaintext
ctx_ddl.create_section_group('basicgroup', 'BASIC_SECTION_GROUP');
ctx_ddl.add_field_section('basicgroup', 'Author', 'A', FALSE);
end;
```

Because the Author field section is not visible, to find text within the Author section, you must use the WITHIN operator as follows:

' (Martin Luther King) WITHIN Author'

A query of Martin Luther King without the WITHIN operator does not return instances of this term in field sections. To query text within field sections without specifying WITHIN, you must set the visible flag to TRUE when you create the section as follows:

```plaintext
ctx_ddl.add_field_section('basicgroup', 'Author', 'A', TRUE);
end;
```

Creating Sections for <META> Tags
When you use the HTML_SECTION_GROUP, you can create sections for META tags.

Consider an HTML document that has a META tag as follows:

```html
<META NAME="author" CONTENT="ken">
```

To create a field section that indexes the CONTENT attribute for the <META NAME="author"> tag:

```plaintext
ctx_ddl.create_section_group('myhtmlgroup', 'HTML_SECTION_GROUP');
ctx_ddl.add_field_section('myhtmlgroup', 'author', 'META@AUTHOR');
end
```

After indexing with section group mygroup, query the document as follows:

'ken WITHIN author'

Limitations

Nested Sections
Field sections cannot be nested. For example, if you define a field section to start with <TITLE> and define another field section to start with <FOO>, the two sections cannot be nested as follows:

```html
<TITLE> dog <FOO> cat </FOO> </TITLE>
```

To work with nested section define them as zone sections.
Replaced Sections

Repeated field sections are allowed, but WITHIN queries treat them as a single section. The following is an example of repeated field section in a document:

<TITLE> cat </TITLE>
<TITLE> dog </TITLE>

The query (dog and cat) within title returns the document, even though these words occur in different sections.

To have WITHIN queries distinguish repeated sections, define them as zone sections.

Related Topics

WITHIN operator on page 3-60 in Chapter 3, "Oracle Text CONTAINS Query Operators".
Section Group Types on page 2-77 in Chapter 2, "Oracle Text Indexing Elements".
CREATE_SECTION_GROUP on page 7-38
ADD_ZONE_SECTION on page 7-28
ADD_SPECIAL_SECTION on page 7-18
REMOVE_SECTION on page 7-68
DROP_SECTION_GROUP on page 7-48
**ADD_INDEX**

Use this procedure to add a sub-index to a catalog index preference. Create this preference by naming one or more columns in the base table.

Because you create sub-indexes to improve the response time of structured queries, the column you add should be used in the `structured_query` clause of the `CATSEARCH` operator at query time.

**Syntax**

```sql
CTX_DDL.ADD_INDEX(set_name in varchar2,
column_list varchar2,
storage_clause varchar2);
```

**set_name**
Specify the name of the index set.

**column_list**
Specify a comma separated list of columns to index. At index time, any column listed here cannot have a NULL value in any row in the base table. If any row is NULL during indexing, then an error is raised.

Always ensure that your columns have non-NULL values before and after indexing.

**storage_clause**
Specify a storage clause.

**Example**

Consider a table called `AUCTION` with the following schema:

```sql
create table auction(
  item_id number,
  title varchar2(100),
  category_id number,
  price number,
  bid_close date);
```

Assume that queries on the table involve a mandatory text query clause and optional structured conditions on `category_id`. Results must be sorted based on `bid_close`.

You can create a catalog index to support the different types of structured queries a user might enter.

To create the indexes, first create the index set preference then add the required indexes to it:

```sql
begin
  ctx_ddl.create_index_set('auction_iset');
  ctx_ddl.add_index('auction_iset','bid_close');
  ctx_ddl.add_index('auction_iset','category_id, bid_close');
end;
```

Create the combined catalog index with `CREATE INDEX` as follows:

```sql
create index auction_titlex on AUCTION(title) indextype is CTXCAT parameters ('index set auction_iset');
```
Querying

To query the title column for the word `pokemon`, enter regular and mixed queries as follows:

```sql
select * from AUCTION where CATSEARCH(title, 'pokemon', NULL) > 0;
select * from AUCTION where CATSEARCH(title, 'pokemon', 'category_id=99 order by bid_close desc') > 0;
```

Notes

VARCHAR2 columns in the column list of a CTXCAT index of an index set cannot exceed 30 bytes.

Related Topic

"REMOVE_INDEX" on page 7-66.
Use this procedure to change the metadata of a document that has been specified as an MDATA section. After this call, MDATA queries involving the named MDATA value will find documents with the given MDATA value.

There are two versions of CTX_DDL.ADD_MDATA: one for adding a single metadata value to a single rowid, and one for handling multiple values, multiple rowids, or both.

CTX_DDL.ADD_MDATA is transactional; it takes effect immediately in the calling session, can be seen only in the calling session, can be reversed with a ROLLBACK command, and must be committed to take permanent effect.

Use CTX_DDL.REMOVE_MDATA to remove metadata values from already-indexed documents. Only the owner of the index is allowed to call ADD_MDATA and REMOVE_MDATA.

Syntax

This is the syntax for adding a single value to a single rowid:

```
CTX_DDL.ADD_MDATA(
    idx_name           IN VARCHAR2,
    section_name       IN VARCHAR2,
    mdata_value        IN VARCHAR2,
    mdata_rowid        IN VARCHAR2,
    [part_name]        IN VARCHAR2
);
```

idx_name
Name of the text index that contains the named rowid.

section_name
Name of the MDATA section.

mdata_value
The metadata value to add to the document.

mdata_rowid
The rowid to which to add the metadata value.

[part_name]
Name of the index partition, if any. Must be provided for local partitioned indexes and must be NULL for global, non-partitioned indexes.

This is the syntax for handling multiple values, multiple rowids, or both. This version is more efficient for large numbers of new values or rowids.

```
CTX_DDL.ADD_MDATA(
    idx_name           IN VARCHAR2,
    section_name       IN VARCHAR2,
    mdata_values       SYS.ODCIVARCHAR2LIST,
    mdata_rowids       SYS.ODCIRIDLIST,
    [part_name]        IN VARCHAR2
);
```

idx_name
Name of the text index that contains the named rowids.
section_name
Name of the MDATA section.

mdata_values
List of metadata values. If a metadata value contains a comma, the comma must be escaped with a backslash.

mdata_rowids
The rowids to which to add the metadata values.

[part_name]
Name of the index partition, if any. Must be provided for local partitioned indexes and must be NULL for global, non-partitioned indexes.

Example
This example updates a single value:

```sql
select rowid from mytab where contains(text, 'MDATA(sec, value>')>0;
No rows returned
exec ctx_ddl.add_mdata('my_index', 'sec', 'value', 'ABC');
select rowid from mytab where contains(text, 'MDATA(sec, value>')>0;
ROWID
-----
ABC
```

This example updates multiple values:

```sql
begin
ctx_ddl.add_mdata('my_index', 'sec',
    sys.odcivarchar2list('value1','value2','value3'),
    sys.odciridlist('ABC','DEF'));
end;
```

This is equivalent to:

```sql
begin
ctx_ddl.add_mdata('my_index', 'sec', 'value1', 'ABC');
ctx_ddl.add_mdata('my_index', 'sec', 'value1', 'DEF');
ctx_ddl.add_mdata('my_index', 'sec', 'value2', 'ABC');
ctx_ddl.add_mdata('my_index', 'sec', 'value2', 'DEF');
ctx_ddl.add_mdata('my_index', 'sec', 'value3', 'ABC');
ctx_ddl.add_mdata('my_index', 'sec', 'value3', 'DEF');
end;
```

Notes
If a rowid is not yet indexed, CTX_DDL.ADD.MDATA completes without error, but an error is logged in CTX_USER_INDEX_ERRORS.

These updates are updates directly on the index itself, not on the actual contents stored in the base table. Therefore, they will not survive when the Text index is rebuilt.

Related Topics
See also "ADD_MDATA_SECTION" on page 7-12; "REMOVE_MDATA" on page 7-67; "MDATA" on page 3-28; as well as the Section Searching chapter of the Oracle Text Application Developer’s Guide.
ADD_MDATA_COLUMN

Use this procedure to map the FILTER BY column named in column_name to the MDATA section named in section_name.

Syntax

The syntax is as follows:

```sql
CTX_DDL.ADD_MDATA_COLUMN(
    group_name IN VARCHAR2,
    section_name IN VARCHAR2,
    column_name IN VARCHAR2,
);
```

- **group_name**
  Name of the group that contains the section.

- **section_name**
  Name of the MDATA section.

- **column_name**
  Name of the FILTER BY column to add to the MDATA section.

Restrictions

MDATA sections that are created with CTX_DDL.ADD_MDATA_COLUMN cannot have their values changed using CTX_DDL.ADD_MDATA or CTX_DDL.REMOVE_MDATA. Doing so will result in errors being returned. The section values must be updated using SQL.

Notes

- The stored datatype for MDATA sections is text. Therefore, the value of the FILTER BY column is converted to text during indexing. For non-text datatypes, the FILTER BY columns are normalized to an internal format during indexing. If the section is queried with an MDATA operator, then the MDATA query string will also be normalized to the internal format before processing.

- When a FILTER BY column is mapped as MDATA, the cost-based optimizer in Oracle Text tries to avoid using the Oracle Text composite domain index to process range predicate(s) on that FILTER BY column. This is because range predicates on MDATA FILTER BY columns are processed less efficiently than if they were declared as SDATA. For this reason, you should not add a FILTER BY column as MDATA if you plan to do range searches on the column.

Related Topics

- MDATA on page 3-28
- ADD_MDATA_SECTION on page 7-12
- REMOVE_MDATA on page 7-67
- ADD_SDATA_COLUMN on page 7-14

See Also:

Chapter 8, "Searching Document Sections in Oracle Text" in Oracle Text Application Developer’s Guide
ADD_MDATA_SECTION

Use this procedure to add an MDATA section, with an accompanying value, to an existing section group. MDATA sections cannot be added to Null Section groups, Path Section groups, or Auto Section groups.

Section values undergo a simplified normalization:

- Leading and trailing whitespace on the value is removed.
- The value is truncated to 64 bytes.
- The value is converted to upper case.
- The value is indexed as a single value; if the value consists of multiple words, it is not broken up.
- Case is preserved. If the document is dynamically generated, then implement case-insensitivity by uppercasing MDATA values and making sure to search only in uppercase.

Use CTX_DDL.REMOVE_SECTION to remove sections.

Syntax

```
CTX_DDL.ADD_MDATA_SECTION(
    group_name    IN VARCHAR2,
    section_name  IN VARCHAR2,
    tag           IN VARCHAR2,
);
```

`group_name`
Name of the section group that will contain the MDATA section.

`section_name`
Name of the MDATA section.

`tag`
The value of the MDATA section. For example, if the section is `<AUTHOR>`, the value could be Cynthia Kadohata (author of the novel The Floating World). More than one tag can be assigned to a given MDATA section.

Example

This example creates an MDATA section called auth.

```
ctx_ddl.create_section_group('htmgp', 'HTML_SECTION_GROUP');
ctx_ddl.add_mdata_section('htmgp', 'auth', 'author');
```

Related Topics

- ADD_MDATA on page 7-9
- REMOVE_MDATA on page 7-67
- MDATA on page 3-28 in Chapter 3, "Oracle Text CONTAINS Query Operators"
- CREATE_SECTION_GROUP on page 7-38 of this chapter

The Section Searching chapter of the Oracle Text Application Developer’s Guide
ADD_NDATA_SECTION

Use this procedure to find matches that are spelled in a similar way. The value of an NDATA section is extracted from the document text like other sections, but is indexed as name data. NDATA sections are stored in the CTX_USER_SECTIONS view.

Syntax

```sql
CTX_DDL.ADD_NDATA_SECTION (
    group_name    IN VARCHAR2,
    section_name  IN VARCHAR2,
    tag           IN VARCHAR2);
```

- `group_name` Name of the group that contains the section.
- `section_name` Name of the NDATA section.
- `tag` Name of the tag that marks the start of a section. For example, if the tag is `<H1>`, specify H1. The start tag you specify must be unique within a section group.

Notes

NDATA sections support both single and multi-byte data, however, there are character- and term-based limitations. NDATA section data that is indexed is constrained as follows:

- number of characters in a single, white space delimited term
  511
- number of white space delimited terms
  255
- total number of characters, including white spaces
  511

NDATA section data that exceeds these constraints are truncated.

Example

The following example defines a section group `namegroup` of the BASIC_SECTION_GROUP type. It then creates an NDATA section in `namegroup` called `firstname`.

```sql
begin
    ctx_ddl.create_section_group('namegroup', 'BASIC_SECTION_GROUP');
    ctx_ddl.add_ndata_section('namegroup', 'firstname', 'fname1');
end;
```
ADD_SDATA_COLUMN

Use this procedure to map the FILTER BY or ORDER BY column named in column_name to the SDATA section named in section_name. By default, all FILTER BY columns are mapped as SDATA.

Syntax

The syntax is as follows:

```sql
CTX_DDL.ADD_SDATA_COLUMN(
  group_name         IN VARCHAR2,
  section_name       IN VARCHAR2,
  column_name        IN VARCHAR2,
);
```

**group_name**
Name of the group that contains the section.

**section_name**
Name of the SDATA section.

**column_name**
Name of the FILTER BY column to add to the SDATA section.

Notes

- Mapping FILTER BY columns to sections is optional. If no section mapping exists for a FILTER BY column, then it is mapped to an SDATA section, and the section name will be the name of the FILTER BY column.

- If a section group is not specified during CREATE INDEX of a composite domain index, then system default section group settings will be used, and a SDATA section will be created for each of the FILTER BY and ORDER BY columns.

---

**Note:** Because section name does not allow certain special characters and is case insensitive, if the column name is case sensitive or contains special characters, then an error will be raised. To work around this problem, you need to map the column to an MDATA or SDATA section before creating the index. Refer to CTX_DDL.ADD_MDATA_COLUMN or CTX_DDL.ADD_SDATA_COLUMN in this chapter.

---

- An error will be raised if a column mapped to MDATA also appears in the ORDER BY column clause.

- Column section names are unique to their section group. That is, you cannot have an MDATA column section named FOO if you already have an MDATA column section named FOO. Furthermore, you cannot have a field section named FOO if you already have an SDATA column section named FOO. This is true whether it is implicitly created (by CREATE INDEX for FILTER BY or ORDER BY clauses) or explicitly created (by CTX_DDL.ADD_SDATA_COLUMN).

- One section name can only be mapped to one FILTER BY column, and vice versa. For example, mapping a section to more than one column or mapping a column to more than one section is not allowed.
■ Column sections can be added to any type of section group, including the NULL section group.

Related Topics

SDATA on page 3-44 in Chapter 3, "Oracle Text CONTAINS Query Operators"
ADD_SDATA_SECTION on page 7-16

See Also: Chapter 8, "Searching Document Sections in Oracle Text"
in Oracle Text Application Developer's Guide
ADD_SDATA_SECTION

This procedure adds an SDATA section to a section group. By default, all FILTER BY columns are mapped as SDATA.

Syntax

The syntax is as follows:

```
CTX_DDL.ADD_SDATA_SECTION(
    group_name         IN VARCHAR2,
    section_name       IN VARCHAR2,
    tag                IN VARCHAR2,
    datatype           IN VARCHAR2, default NULL,
);
```

**group_name**
Name of the group that contains the section.

**section_name**
Name of the SDATA section.

**tag**
Name of the tag to add to the SDATA section.

**datatype**
Specifies the stored format for the data, as well as the semantics of comparison in later use in SDATA operators. The default is VARCHAR2, but if specified must be one of the following values:

- VARCHAR2
- CHAR
- RAW
- NUMBER
- DATE

The VARCHAR2 datatype stores up to 249 bytes of character data in the database character set. Values larger than this result in a per-document indexing error. Note that leading and trailing whitespace are always trimmed from SDATA section values when extracted by the sectioner. This is different than SDATA columns. Column values are never trimmed. No lexing is performed on the value from either kind of SDATA.

The CHAR datatype stores up to 249 bytes of character data in the database character set. Values larger than this result in a per-document indexing error. Note that leading and trailing whitespace are always trimmed from SDATA section values when extracted by the sectioner. This is different than SDATA columns. Column values are never trimmed. No lexing is performed on the value from either kind of SDATA. To be consistent with SQL, the comparisons of CHAR datatype SDATA values are blank-padded comparisons.

RAW datatype stores up to 249 bytes of binary data. Values larger than this result in a per-document indexing error. The value is converted from hexadecimal string representation. That is, to store a value of 65, the document should look like `<TAG>40</TAG>`, and not `<TAG>65</TAG>` or `<TAG>A</TAG>`.
The **DATE** datatype values must conform to the following format: `YYYY-MM-DD` or `YYYY-MM-DD HH24:MI:SS`. That is, to store a **DATE** value of "Nov. 24, 2006 10:32pm 36sec", the document should look like `<TAG>2006-11-24 22:32:36</TAG>`.

**Limitations**

- **SDATA** are single-occurrence only. If multiple instances of an **SDATA** tag are encountered in a single document, then later instances supersede the value set by earlier instances. This means that the last occurrence of an **SDATA** tag takes effect.
- If no **SDATA** tag occurs in a given document, then this is treated as an **SDATA** value of **NULL**.
- Empty **SDATA** tags are treated as **NULL** values.
- **SDATA** sections cannot be nested. Sections that are nested inside are ignored.

**Related Topics**

- "**SDATA**" on page 3-44 in Chapter 3, "Oracle Text CONTAINS Query Operators"
- "**ADD_SDATA_COLUMN**" on page 7-14 of this chapter

**See Also:** Chapter 8, "Searching Document Sections in Oracle Text" in *Oracle Text Application Developer’s Guide*
ADD_SPECIAL_SECTION

Adds a special section, either SENTENCE or PARAGRAPH, to a section group. This enables searching within sentences or paragraphs in documents with the WITHIN operator.

A special section in a document is a section which is not explicitly tagged like zone and field sections. The start and end of special sections are detected when the index is created. Oracle Text supports two such sections: paragraph and sentence.

The sentence and paragraph boundaries are determined by the lexer. For example, the lexer recognizes sentence and paragraph section boundaries as follows:

<table>
<thead>
<tr>
<th>Special Section</th>
<th>Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>SENTENCE</td>
<td>WORD/PUNCT/WHITESPACE</td>
</tr>
<tr>
<td>PARAGRAPH</td>
<td>WORD/PUNCT/NEWLINE/WHITESPACE (indented paragraph)</td>
</tr>
<tr>
<td></td>
<td>WORD/PUNCT/NEWLINE/NEWLINE (block paragraph)</td>
</tr>
</tbody>
</table>

The punctuation, whitespace, and newline characters are determined by your lexer settings and can be changed.

If the lexer cannot recognize the boundaries, no sentence or paragraph sections are indexed.

Syntax

```
CTX_DDL.ADD_SPECIAL_SECTION(
    group_name IN VARCHAR2,
    section_name IN VARCHAR2);
```

- **group_name**
  Specify the name of the section group.

- **section_name**
  Specify SENTENCE or PARAGRAPH.

Example

The following example enables searching within sentences within HTML documents:

```
begin
ctx_ddl.create_section_group('htmgroup', 'HTML_SECTION_GROUP');
ctx_ddl.add_special_section('htmgroup', 'SENTENCE');
end;
```

Add zone sections to the group to enable zone searching in addition to sentence searching. The following example adds the zone section Headline to the section group htmgroup:

```
begin
ctx_ddl.create_section_group('htmgroup', 'HTML_SECTION_GROUP');
ctx_ddl.add_special_section('htmgroup', 'SENTENCE');
ctx_ddl.add_zone_section('htmgroup', 'Headline', 'H1');
end;
```
If you are only interested in sentence or paragraph searching within documents and not interested in defining zone or field sections, then use the NULL_SECTION_GROUP as follows:

```java
begin
  ctx_ddl.create_section_group('nullgroup', 'NULL_SECTION_GROUP');
  ctx_ddl.add_special_section('nullgroup', 'SENTENCE');
end;
```

Related Topics

- `WITHIN` operator in Chapter 3, "Oracle Text CONTAINS Query Operators".
- "Section Group Types" in Chapter 2, "Oracle Text Indexing Elements".
- `CREATE_SECTION_GROUP`
- `ADD_ZONE_SECTION`
- `ADD_FIELD_SECTION`
- `REMOVE_SECTION`
- `DROP_SECTION_GROUP`
ADD_STOPCLASS

Adds a stopclass to a stoplist. A stopclass is a class of tokens that is not to be indexed.

Syntax

```sql
CTX_DDL.ADD_STOPCLASS(
    stoplist_name  in   varchar2,
    stopclass      in   varchar2
);
```

**stoplist_name**
Specify the name of the stoplist.

**stopclass**
Specify the stopclass to be added to `stoplist_name`. Currently, only the `NUMBERS` class is supported. It is not possible to create a custom stopclass.

`NUMBERS` includes tokens that follow the number pattern: digits, numgroup, and numjoin only. Therefore, `123ABC` is not a number, nor is `A123`. These are labeled as `MIXED`. `$123` is not a number (this token is not common in a text index because non-alphanumerics become whitespace by default). In the United States, `123.45` is a number, but `123.456.789` is not; in Europe, where numgroup may be `',` the reverse is true.

The maximum number of stopwords, stoptemes, and stopclasses you can add to a stoplist is 4095.

**Example**

The following example adds a stopclass of `NUMBERS` to the stoplist `mystop`:

```sql
begin
    ctx_ddl.add_stopclass('mystop', 'NUMBERS');
end;
```

**Related Topics**

CREATE_STOPLIST
REMOVE_STOPCLASS
DROP_STOPLIST
**ADD_STOP_SECTION**

Adds a stop section to an automatic section group. Adding a stop section causes the automatic section indexing operation to ignore the specified section in XML documents.

---

**Note:** Adding a stop section causes no section information to be created in the index. However, the text within a stop section is always searchable.

---

Adding a stop section is useful when your documents contain many low information tags. Adding stop sections also improves indexing performance with the automatic section group.

The number of stop sections you can add is unlimited.

Stop sections do not have section names and hence are not recorded in the section views.

**Syntax**

```plaintext
CTX_DDL.ADD_STOP_SECTION(
    section_group IN VARCHAR2,
    tag  IN VARCHAR2);
```

**section_group**

Specify the name of the automatic section group. If you do not specify an automatic section group, then this procedure returns an error.

**tag**

Specify the tag to ignore during indexing. This parameter is case-sensitive. Defining a stop tag as such also stops the tag's attribute sections, if any.

Qualify the tag with document type in the form `(doctype)tag`. For example, if you wanted to make the `<fluff>` tag a stop section only within the `mydoc` document type, specify `(mydoc)fluff` for `tag`.

**Example**

**Defining Stop Sections**

The following example adds a stop section identified by the tag `<fluff>` to the automatic section group `myauto`:

```plaintext
begin
    ctx_ddl.add_stop_section('myauto', 'fluff');
end;
```

This example also stops any attribute sections contained within `<fluff>`. For example, if a document contained:

```xml
<fluff type='computer'>
```

Then the preceding example also stops the attribute section `fluff@type`. 
Doctype Sensitive Stop Sections

The following example creates a stop section for the tag `<fluff>` only in documents that have a root element of `mydoc`:

```
begin
ctx_ddl.add_stop_section('myauto', '(mydoc)fluff');
end;
```

Related Topics

- [ALTER INDEX](#) in Chapter 1, "Oracle Text SQL Statements and Operators".
- [CREATE_SECTION_GROUP](#)
ADD_STOPTHEME

Adds a single stoptheme to a stoplist. A stoptheme is a theme that is not to be indexed.
In English, query on indexed themes using the ABOUT operator.

Syntax

```sql
CTX_DDL.ADD_STOPTHEME(
    stoplist_name  in   varchar2,
    stoptheme      in   varchar2
);
```

**stoplist_name**
Specify the name of the stoplist.

**stoptheme**
Specify the stoptheme to be added to stoplist_name. The system normalizes the stoptheme you enter using the knowledge base. If the normalized theme is more than one theme, then the system does not process your stoptheme. For this reason, Oracle recommends that you submit single stopthemes.

The maximum number of stopwords, stopthemes, and stopclasses you can add to a stoplist is 4095.

Example

The following example adds the stoptheme banking to the stoplist mystop:

```sql
begin
    ctx_ddl.add_stoptheme('mystop', 'banking');
end;
```

Related Topics

CREATE_STOPLIST
REMOVE_STOPTHEME
DROP_STOPLIST
ABOUT operator in Chapter 3, "Oracle Text CONTAINS Query Operators".
ADD_STOPWORD

Use this procedure to add a single stopword to a stoplist.
To create a list of stopwords, you must call this procedure once for each word.

Syntax

```
CTX_DDL.ADD_STOPWORD(
stoplist_name  in   varchar2,
stopword       in   varchar2,
language       in varchar2 default NULL
);
```

**stoplist_name**
Specify the name of the stoplist.

**stopword**
Specify the stopword to be added.

Language-specific stopwords must be unique across the other stopwords specific to the language. For example, it is valid to have a German *die* and an English *die* in the same stoplist.

The maximum number of stopwords, stopthemes, and stopclasses you can add to a stoplist is 4095.

**language**
Specify the language of stopword when the stoplist you specify with stoplist_name is of type MULTI_STOPLIST. You must specify the globalization support name or abbreviation of an Oracle Text-supported language.

To make a stopword active in multiple languages, specify ALL for this parameter. For example, defining ALL stopwords is useful when you have international documents that contain English fragments that need to be stopped in any language.

An ALL stopword is active in all languages. If you use the multi-lexer, the language-specific lexing of the stopword occurs, just as if it had been added multiple times in multiple specific languages.

Otherwise, specify NULL.

Example

**Single Language Stoplist**
The following example adds the stopwords *because, notwithstanding, nonetheless,* and *therefore* to the stoplist mystop:

```
begin
ctx_ddl.add_stopword('mystop', 'because');
ctx_ddl.add_stopword('mystop', 'notwithstanding');
ctx_ddl.add_stopword('mystop', 'nonetheless');
ctx_ddl.add_stopword('mystop', 'therefore');
end;
```

**Multi-Language Stoplist**
The following example adds the German word *die* to a multi-language stoplist:
begin
  ctx_ddl.add_stopword('mystop', 'Die', 'german');
end;

---

**Note:** Add stopwords after you create the index with ALTER INDEX.

---

**Adding An ALL Stopword**

The following adds the word *the* as an **ALL** stopword to the multi-language stoplist *globallist*:

begin
  ctx_ddl.add_stopword('globallist', 'the', 'ALL');
end;

**Related Topics**

CREATE_STOPLIST
REMOVE_STOPWORD
DROP_STOPLIST
ALTER INDEX in Chapter 1, "Oracle Text SQL Statements and Operators". Appendix E, "Oracle Text Supplied Stoplists"
ADD_SUBLexer

Add a sub-lexer to a multi-lexer preference. A sub-lexer identifies a language in a multi-lexer (multi-language) preference. Use a multi-lexer preference when you want to index more than one language.

Restrictions

The following restrictions apply to using CTX_DDL.ADD_SUBLexer:

- The invoking user must be the owner of the multi-lexer or CTXSYS.
- The lexer_name parameter must name a preference which is a multi-lexer lexer.
- A lexer for default must be defined before the multi-lexer can be used in an index.
- The sub-lexer preference owner must be the same as multi-lexer preference owner.
- The sub-lexer preference must not be a multi-lexer lexer.
- A sub-lexer preference cannot be dropped while it is being used in a multi-lexer preference.
- CTX_DDL.ADD_SUBLexer records only a reference. The sub-lexer values are copied at create index time to index value storage.

Syntax

CTX_DDL.ADD_SUBLexer(
   lexer_name in varchar2,
   language  in varchar2,
   sub_lexer in varchar2,
   alt_value in varchar2 default null
);

lexer_name
Specify the name of the multi-lexer preference.

language
Specify the globalization support language name or abbreviation of the sub-lexer. For example, specify JAPANESE or JA for Japanese.

The sub-lexer you specify with sub_lexer is used when the language column has a value case-insensitive equal to the globalization support name of abbreviation of language.

Specify DEFAULT to assign a default sub-lexer to use when the value of the language column in the base table is null, invalid, or unmapped to a sub-lexer. The DEFAULT lexer is also used to parse stopwords.

If a sub-lexer definition for language already exists, then it is replaced by this call.

sub_lexer
Specify the name of the sub-lexer to use for this language.

alt_value
Optionally specify an alternate value for language.

If you specify DEFAULT for language, then you cannot specify an alt_value.
The `alt_value` is limited to 30 bytes and cannot be a globalization support language name, abbreviation, or DEFAULT.

**Example**

This example shows how to create a multi-language text table and how to set up the multi-lexer to index the table.

Create the multi-language table with a primary key, a text column, and a language column as follows:

```sql
create table globaldoc (
    doc_id number primary key,
    lang varchar2(3),
    text clob
);
```

Assume that the table holds mostly English documents, with an occasional German or Japanese document. To handle the three languages, you must create three sub-lexers: one for English, one for German, and one for Japanese as follows:

```sql
ctx_ddl.create_preference('english_lexer','basic_lexer');
ctx_ddl.set_attribute('english_lexer','index_themes','yes');
ctx_ddl.set_attribute('english_lexer','theme_language','english');

ctx_ddl.create_preference('german_lexer','basic_lexer');
ctx_ddl.set_attribute('german_lexer','composite','german');
ctx_ddl.set_attribute('german_lexer','mixed_case','yes');
ctx_ddl.set_attribute('german_lexer','alternate_spelling','german');

ctx_ddl.create_preference('japanese_lexer','japanese_vgram_lexer');
```

Create the multi-lexer preference:

```sql
ctx_ddl.create_preference('global_lexer', 'multi_lexer');
```

Because the stored documents are mostly English, make the English lexer the default:

```sql
ctx_ddl.add_sub_lexer('global_lexer','default','english_lexer');
```

Add the German and Japanese lexers in their respective languages. Also assume that the language column is expressed in ISO 639-2, so add those as alternative values.

```sql
ctx_ddl.add_sub_lexer('global_lexer','german','german_lexer','ger');
ctx_ddl.add_sub_lexer('global_lexer','japanese','japanese_lexer','jpn');
```

Create the index `globalx`, specifying the multi-lexer preference and the language column in the parameters string as follows:

```sql
create index globalx on globaldoc(text) indextype is ctxsys.context
parameters ('lexer global_lexer language column lang');
```
ADD_ZONE_SECTION

Creates a zone section and adds the section to an existing section group. This enables zone section searching with the WITHIN operator.

Zone sections are sections delimited by start and end tags. The <B> and </B> tags in HTML, for instance, marks a range of words which are to be rendered in boldface.

Zone sections can be nested within one another, can overlap, and can occur more than once in a document.

Syntax

```sql
CTX_DDL.ADD_ZONE_SECTION(
    group_name     in    varchar2,
    section_name   in    varchar2,
    tag            in    varchar2
);
```

- **group_name**: Specify the name of the section group to which `section_name` is added.

- **section_name**: Specify the name of the section to add to the `group_name`. Use this name to identify the section in WITHIN queries. Avoid using names that contain non-alphanumeric characters such as _, because most of these characters are special must be escaped in queries. Section names are case-insensitive.

Within the same group, zone section names and field section names cannot be the same. The terms Paragraph and Sentence are reserved for special sections.

Section names need not be unique across tags. You can assign the same section name to more than one tag, making details transparent to searches.

- **tag**: Specify the pattern which marks the start of a section. For example, if <H1> is the HTML tag, specify H1 for `tag`. The start tag you specify must be unique within a section group.

Oracle Text knows what the end tags look like from the `group_type` parameter you specify when you create the section group.

If `group_name` is an HTML_SECTION_GROUP, you can create zone sections for the META tag’s NAME/CONTENT attribute pairs. To do so, specify `tag` as meta@namevalue where namevalue is the value of the NAME attribute whose CONTENT attributes are to be indexed as a section. Refer to the example.

If `group_name` is an XML_SECTION_GROUP, you can optionally qualify `tag` with a document type (root element) in the form (doctype)tag. Doing so makes `section_name` sensitive to the XML document type declaration. Refer to the example.

Examples

**Creating HTML Sections**

The following example defines a section group called htmgroup of type HTML_SECTION_GROUP. It then creates a zone section in htmgroup called headline identified by the <H1> tag:
begin
cxt_ddl.create_section_group('htmgroup', 'HTML_SECTION_GROUP');
cxt_ddl.add_zone_section('htmgroup', 'heading', 'H1');
end;

After indexing with section group htmgroup, query within the heading section by issuing a query as follows:

`Oracle WITHIN heading`

**Creating Sections for `<META NAME>` Tags**

You can create zone sections for HTML META tags when you use the `HTML_SECTION_GROUP`.

Consider an HTML document that has a META tag as follows:

```html
<META NAME="author" CONTENT="ken">
```

To create a zone section that indexes all CONTENT attributes for the META tag whose NAME value is author:

begin

cxt_ddl.create_section_group('htmgroup', 'HTML_SECTION_GROUP');
cxt_ddl.add_zone_section('htmgroup', 'author', 'meta@author');
end

After indexing with section group htmgroup, query the document as follows:

`ken WITHIN author`

**Creating Document Type Sensitive Sections (XML Documents Only)**

You have an XML document set that contains the `<book>` tag declared for different document types (DTDs). You want to create a distinct book section for each document type.

Assume that `myDTDname` is declared as an XML document type as follows:

```xml
<!DOCTYPE myDTDname>
<myDTDname>
...
```

(Note: the DOCTYPE must match the top-level tag.)

Within `myDTDname`, the element `<book>` is declared. For this tag, create a section named mybooksec that is sensitive to the tag's document type as follows:

begin

cxt_ddl.create_section_group('myxmlgroup', 'XML_SECTION_GROUP');
cxt_ddl.add_zone_section('myxmlgroup', 'mybooksec', '(myDTDname)book');
end;

**Notes**

**Repeated Sections**

Zone sections can repeat. Each occurrence is treated as a separate section. For example, if `<H1>` denotes a heading section, they can repeat in the same documents as follows:

```
<H1> The Brown Fox </H1>
<br/>
<H1> The Gray Wolf </H1>
```
Assuming that these zone sections are named Heading, the query *Brown* WITHIN *Heading* returns this document. However, a query of *(Brown and Gray)* WITHIN *Heading* does not.

**Overlapping Sections**
Zone sections can overlap each other. For example, if `<B>` and `<I>` denote two different zone sections, they can overlap in document as follows:

plain <B> bold <I> bold and italic </B> only italic </I> plain

**Nested Sections**
Zone sections can nest, including themselves as follows:

```html
<TD> <TABLE><TD>nested cell</TD></TABLE></TD>
```

Using the WITHIN operator, you can write queries to search for text in sections within sections. For example, assume the `BOOK1`, `BOOK2`, and `AUTHOR` zone sections occur as follows in documents `doc1` and `doc2`:

**doc1:**

```html
<book1> <author>Scott Tiger</author> This is a cool book to read.</book1>
```

**doc2:**

```html
<book2> <author>Scott Tiger</author> This is a great book to read.</book2>
```

Consider the nested query:

`'(Scott within author) within book1'`

This query returns only `doc1`.

**Related Topics**

WITHIN operator in Chapter 3, "Oracle Text CONTAINS Query Operators"
Section Group Types in Chapter 2, "Oracle Text Indexing Elements"
CREATE_SECTION_GROUP
ADD_FIELD_SECTION
ADD_SPECIAL_SECTION
REMOVE_SECTION
DROP_SECTION_GROUP
COPY_POLICY

Creates a new policy from an existing policy or index.

Syntax

```sql
ctx_ddl.copy_policy(
    source_policy VARCHAR2,
    policy_name   VARCHAR2
);
```

**source_policy**
The name of the policy or index being copied.

**policy_name**
The name of the new policy copy.

The preference values are copied from the `source_policy`. Both the source policy or index and the new policy must be owned by the same database user.
CREATE_INDEX_SET

CREATE_INDEX_SET

Creates an index set for CTXCAT index types. Name this index set in the parameter clause of CREATE INDEX when you create a CTXCAT index.

Syntax

CTX_DDL.CREATE_INDEX_SET(set_name in varchar2);

set_name
Specify the name of the index set. Name this index set in the parameter clause of CREATE INDEX when you create a CTXCAT index.
CREATE_POLICY

Creates a policy to use with the CTX_DOC.POLICY_* procedures and the ORA:CONTAINS function. ORA:CONTAINS is a function you use within an XPATH query expression with existsNode().

See Also: Oracle XML DB Developer’s Guide

Syntax

```
CTX_DDL.CREATE_POLICY(
    policy_name     IN VARCHAR2,
    filter          IN VARCHAR2 DEFAULT NULL,
    section_group   IN VARCHAR2 DEFAULT NULL,
    lexer           IN VARCHAR2 DEFAULT NULL,
    stoplist        IN VARCHAR2 DEFAULT NULL,
    wordlist        IN VARCHAR2 DEFAULT NULL);
```

**policy_name**
Specify the name for the new policy. Policy names and Text indexes share the same namespace.

**filter**
Specify the filter preference to use.

**section_group**
Specify the section group to use. You can specify any section group that is supported by CONTEXT index.

**lexer**
Specify the lexer preference to use. Your INDEX_THEMES attribute must be disabled.

**stoplist**
Specify the stoplist to use.

**wordlist**
Specify the wordlist to use.

Example

Create mylex lexer preference named mylex.

```
begin
    ctx_ddl.create_preference('mylex', 'BASIC_LEXER');
    ctx_ddl.set_attribute('mylex', 'printjoins', '_-');
    ctx_ddl.set_attribute('mylex', 'index_themes', 'NO');
    ctx_ddl.set_attribute('mylex', 'index_text', 'YES');
end;
```

Create a stoplist preference named mystop.

```
begin
    ctx_ddl.create_stoplist('mystop', 'BASIC_STOPLIST');
    ctx_ddl.add_stopword('mystop', 'because');
    ctx_ddl.add_stopword('mystop', 'nonetheless');
    ctx_ddl.add_stopword('mystop', 'therefore');
end;
```
Create a wordlist preference named 'mywordlist'.

```sql
begin
  ctx_ddl.create_preference('mywordlist', 'BASIC_WORDLIST');
  ctx_ddl.set_attribute('mywordlist', 'FUZZY_MATCH', 'ENGLISH');
  ctx_ddl.set_attribute('mywordlist', 'FUZZY_SCORE', '0');
  ctx_ddl.set_attribute('mywordlist', 'FUZZY_NUMRESULTS', '5000');
  ctx_ddl.set_attribute('mywordlist', 'SUBSTRING_INDEX', 'TRUE');
  ctx_ddl.set_attribute('mywordlist', 'STEMMER', 'ENGLISH');
  end;
```

```sql
exec ctx_ddl.create_policy('my_policy', NULL, NULL, 'mylex', 'mystop', 'mywordlist');
```

or

```sql
exec ctx_ddl.create_policy(policy_name => 'my_policy',
                          lexer => 'mylex',
                          stoplist => 'mystop',
                          wordlist => 'mywordlist');
```

Then enter the following `existsNode()` query with your own defined policy:

```sql
select id from xmltab
  where existsNode(doc, '/book/chapter[ ora:contains(summary,"dog or cat",
      "my_policy") >0 ]', 'xmlns:ora="http://xmlns.oracle.com/xdb" ')=1;
```

Update the policy with the following:

```sql
exec ctx_ddl.update_policy(policy_name => 'my_policy', lexer => 'my_new_lex');
```

Drop the policy with the following:

```sql
exec ctx_ddl.drop_policy(policy_name => 'my_policy');
```
CREATE_PREFERENCE

Creates a preference in the Text data dictionary. Specify preferences in the parameter string of CREATE_INDEX or ALTER_INDEX.

Caution: CTX_DDL.CREATE_PREFERENCE does not respect the current schema as set by ALTER SESSION SET current_schema. Therefore, if you need to create or delete a preference owned by another user, then you must explicitly state this, and you must have the CREATE ANY TABLE system privilege.

The WebIV note 249991.1 discusses this issue further.

Workaround: Change dbms_xdbt to prepend the index name with XDB as follows:

```sql
indexname CONSTANT VARCHAR2(32) := 'XDB.XDB$CI'
```

Where `indexname` is the name of the CONTEXT index.

Syntax

```sql
CTX_DDL.CREATE_PREFERENCE(preference_name  in varchar2,
object_name      in varchar2);
```

- **preference_name**
  Specify the name of the preference to be created.

- **object_name**
  Specify the name of the preference type.

See Also: For a complete list of preference types and their associated attributes, see Chapter 2, "Oracle Text Indexing Elements"

Examples

Creating Text-only Index

The following example creates a lexer preference that specifies a text-only index. It does so by creating a BASICLexer preference called `my Lexer` with CTX_DDL.CREATE_PREFERENCE. It then calls CTX_DDL.SET_ATTRIBUTE twice, first specifying YES for the INDEX_TEXT attribute, then specifying NO for the INDEX_THEMES attribute.

```sql
begin
  ctx_ddl.create_preference('my Lexer', 'BASICLexer');
  ctx_ddl.set_attribute('my Lexer', 'INDEX_TEXT', 'YES');
  ctx_ddl.set_attribute('my Lexer', 'INDEX_THEMES', 'NO');
end;
```

Specifying File Data Storage

The following example creates a data storage preference called `mypref` that tells the system that the files to be indexed are stored in the operating system. The example then uses CTX_DDL.SET_ATTRIBUTE to set the PATH attribute of to the directory `/docs`. 
begin
ctx_ddl.create_preference('mypref', 'FILE_DATASTORE');
ctx_ddl.set_attribute('mypref', 'PATH', '/docs');
end;

See Also: For more information about data storage, see Datastore Types in Chapter 2, "Oracle Text Indexing Elements"

Creating Master/Detail Relationship
Use CTX_DDL.CREATE_PREFERENCE to create a preference with DETAIL_DATASTORE. Use CTX_DDL.SET_ATTRIBUTE to set the attributes for this preference. The following example shows how this is done:

begin
ctx_ddl.create_preference('my_detail_pref', 'DETAIL_DATASTORE');
ctx_ddl.set_attribute('my_detail_pref', 'binary', 'true');
ctx_ddl.set_attribute('my_detail_pref', 'detail_table', 'my_detail');
ctx_ddl.set_attribute('my_detail_pref', 'detail_key', 'article_id');
ctx_ddl.set_attribute('my_detail_pref', 'detail_lineno', 'seq');
ctx_ddl.set_attribute('my_detail_pref', 'detail_text', 'text');
end;

See Also: For more information about master/detail, see "DETAIL_DATASTORE" in Chapter 2, "Oracle Text Indexing Elements"

Specifying Storage Attributes
The following examples specify that the index tables are to be created in the foo tablespace with an initial extent of 1K:

begin
ctx_ddl.create_preference('mystore', 'BASIC_STORAGE');
ctx_ddl.set_attribute('mystore', 'I_TABLE_CLAUSE',
    'tablespace foo storage (initial 1K)');
ctx_ddl.set_attribute('mystore', 'K_TABLE_CLAUSE',
    'tablespace foo storage (initial 1K)');
ctx_ddl.set_attribute('mystore', 'R_TABLE_CLAUSE',
    'tablespace foo storage (initial 1K)');
ctx_ddl.set_attribute('mystore', 'S_TABLE_CLAUSE',
    'tablespace foo storage (initial 1K)');
ctx_ddl.set_attribute('mystore', 'N_TABLE_CLAUSE',
    'tablespace foo storage (initial 1K)');
ctx_ddl.set_attribute('mystore', 'I_INDEX_CLAUSE',
    'tablespace foo storage (initial 1K)');
end;

See Also: Storage Types in Chapter 2, "Oracle Text Indexing Elements"

Creating Preferences with No Attributes
When you create preferences with types that have no attributes, you need only create the preference, as in the following example which sets the filter to the NULL_FILTER:

begin
ctx_ddl.create_preference('my_null_filter', 'NULL_FILTER');
end;
Notes

If s_table_clause is specified for a storage preference in an index without SDATA, then it has no effect on the index, and the index creation will still succeed.

Related Topics

SET_ATTRIBUTE
DROP_PREFERENCE
CREATE INDEX in Chapter 1, "Oracle Text SQL Statements and Operators"
ALTER INDEX in Chapter 1, "Oracle Text SQL Statements and Operators"
Chapter 2, "Oracle Text Indexing Elements"
CREATE_SECTION_GROUP

Creates a section group for defining sections in a text column.

When you create a section group, you can add to it zone, field, or special sections with ADD_ZONE_SECTION, ADD_FIELD_SECTION, ADD_MDATA_SECTION, or ADD_SPECIAL_SECTION.

When you index, name the section group in the parameter string of CREATE_INDEX or ALTER_INDEX.

After indexing, query within your defined sections with the WITHIN operator.

Syntax

```sql
CTX_DDL.CREATE_SECTION_GROUP(
    group_name in varchar2,
    group_type in varchar2
);
```

**group_name**
Specify the section group name to create as [user.]section_group_name. This parameter must be unique within an owner.

**group_type**
Specify section group type. The group_type parameter can be one of:

<table>
<thead>
<tr>
<th>Section Group Preference</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL_SECTION_GROUP</td>
<td>Use this group type when you define no sections or when you define only SENTENCE or PARAGRAPH sections. This is the default.</td>
</tr>
<tr>
<td>BASIC_SECTION_GROUP</td>
<td>Use this group type for defining sections where the start and end tags are of the form <code>&lt;A&gt;</code> and <code>&lt;/A&gt;</code>. Note: This group type does not support input such as unbalanced parentheses, comments tags, and attributes. Use HTML_SECTION_GROUP for this type of input.</td>
</tr>
<tr>
<td>HTML_SECTION_GROUP</td>
<td>Use this group type for indexing HTML documents and for defining sections in HTML documents.</td>
</tr>
<tr>
<td>XML_SECTION_GROUP</td>
<td>Use this group type for indexing XML documents and for defining sections in XML documents.</td>
</tr>
</tbody>
</table>
CREATE SECTION GROUP

Example

The following command creates a section group called htmgroup with the HTML group type.

```
begin
ctx_ddl.create_section_group('htmgroup', 'HTML_SECTION_GROUP');
end;
```

The following command creates a section group called auto with the AUTO_SECTION_GROUP group type to be used to automatically index tags in XML documents.

```
begin
ctx_ddl.create_section_group('auto', 'AUTO_SECTION_GROUP');
end;
```

Related Topics

- WITHIN operator in Chapter 3, "Oracle Text CONTAINS Query Operators"
- "Section Group Types" in Chapter 2, "Oracle Text Indexing Elements"
- ADD_ZONE_SECTION
- ADD_FIELD_SECTION
- ADD_MDATA_SECTION
- ADD_SPECIAL_SECTION
- REMOVE_SECTION
DROP SECTION_GROUP
CREATE_SHADOW_INDEX

Creates index metadata (or policy) for the specified index. If the index is not partitioned, then it also creates the index tables. This procedure is only supported in Enterprise Edition of Oracle Database.

The following changes are not supported:

- Transition from non-composite domain index to composite, or changing the composite domain index columns.
- Rebuild indexes that have partitioned index tables, for example, $I, $P, $K.

**Note:** For a partitioned index, you must first call this procedure to create the shadow index metadata. This procedure will not create index tables. It has no effect on query, DML, sync, or optimize operations.

**Syntax**

```sql
CTX_DDL.CREATE_SHADOW_INDEX(
    idx_name          IN VARCHAR2,
    parameter_string  IN VARCHAR2 DEFAULT NULL,
    parallel_degree   IN NUMBER, DEFAULT 1
);
```

**idx_name**
The name of a valid CONTEXT indextype.

**parameter_string**
For non-partitioned index, the same string as in `ALTER INDEX`. For partitioned index, the same string as in `ALTER INDEX PARAMETER`.

**parallel_degree**
Reserved for future use. Specify the degree of parallelism. Parallel operation is not currently supported.

**Example**

**Example 7–1 Scheduled Global Index Recreate (Incremental Rebuild)**

In this example, you have the finest control over each stage of `RECREATE_INDEX_ONLINE`. Since `SYNC_INDEX` can take a time limit, you can limit `SYNC_INDEX` during non-business hours and incrementally recreate the index.

```sql
/* create lexer and original index */
exec ctx_ddl.create_preference('us_lexer','basic_lexer');
create index idx on tbl(text) indextype is ctxsys.context
   parameters('lexer us_lexer');

/* create a new lexer */
begin
    ctx_ddl.create_preference('e_lexer','basic_lexer');
    ctx_ddl.set_attribute('e_lexer','base_letter','yes');
    ctx_ddl.create_preference('m_lexer','multi_lexer');
    ctx_ddl.add_sub_lexer('m_lexer','default','us_lexer');
    ctx_ddl.add_sub_lexer('m_lexer','e','e_lexer');
```

Note: For a partitioned index, you must first call this procedure to create the shadow index metadata. This procedure will not create index tables. It has no effect on query, DML, sync, or optimize operations.
CREATE_SHADOW_INDEX

end;
/

/* add new language column to the table for multi-lexer */
alter table tbl add(lang varchar2(10) default 'us');

/* create shadow index */
exec ctx_ddl.create_shadow_index('idx',
    'replace lexer m_lexer language column lang NOPOPULATE');

declar
    idxid integer;
begin
    /* figure out shadow index name */
    select idx_id into idxid from ctx_user_indexes
        where idx_name = 'IDX';
    /* populate pending */
    ctx_ddl.populate_pending('RIO$'||idxid);
    /* time limited sync */
    ctx_ddl.sync_index(idx_name =>'RIO$'||idxid,
        maxtime =>480);
    /* more sync until no pending rows for the shadow index */
end;
/* swap in the shadow index */
exec ctx_ddl.exchange_shadow_index('idx');

Notes

The index name for the shadow index is RIO$index_id. By default it will also populate index tables for non-partitioned indexes, unless NOPOPULATE is specified in CREATE INDEX or in ALTER INDEX. For a local partitioned index, it will only create index metadata without creating the index tables for each partition. Each index can have only one shadow index.

When building a non-partitioned index online, you can first call this procedure to create index metadata and index tables. If you specify POPULATE, then this procedure will populate the index, but will not do swapping. You can schedule the swapping at a later, preferred time.

If you specify NOPOPULATE, it will only create metadata for the index tables, but will not populate them. You must perform POPULATE_PENDING (CTX_DDL.POPULATE_PENDING) to populate the pending queues after running this procedure, and then sync the indexes. This is referred to as incremental recreate.

Queries are all processed normally when this procedure is running.

If POPULATE is specified, then DML is blocked for a very short time at the beginning of populate, after which all further DML is logged into an online pending queue and processed later.

Sync with CTX_DDL.SYNC_INDEX runs normally on the index. OPTIMIZE_INDEX runs without doing anything, but does not return an error.

Related Topics

See also POPULATE | NOPOPULATE in ALTER INDEX and CREATE INDEX in Chapter 1, “Oracle Text SQL Statements and Operators”, and CTX_DDL.DROP_SHADOW_INDEX, CTX_DDL.EXCHANGE_SHADOW_INDEX, CTX_DDL.SYNC_INDEX, and CTX_DDL.POPULATE_PENDING in this chapter.
CREATE_STOPLIST

Use this procedure to create a new, empty stoplist. Stoplists can contain words or themes that are not to be indexed.

You can also create multi-language stoplists to hold language-specific stopwords. A multi-language stoplist is useful when you index a table that contains documents in different languages, such as English, German, and Japanese. When you do so, the text table must contain a language column.

Add either stopwords, stopclasses, or stopthemes to a stoplist using ADD_STOPWORD, ADD_STOPCLASS, or ADD_STOPTHEME. Specify a stoplist in the parameter string of CREATE INDEX or ALTER INDEX to override the default stoplist CTXSYS.DEFAULT_STOPLIST.

Syntax

```sql
CTX_DDL.CREATE_STOPLIST(
    stoplist_name IN VARCHAR2,
    stoplist_type IN VARCHAR2 DEFAULT 'BASIC_STOPLIST');
```

- **stoplist_name**
  Specify the name of the stoplist to be created.

- **stoplist_type**
  Specify BASIC_STOPLIST to create a stoplist for a single language. This is the default.
  Specify MULTI_STOPLIST to create a stoplist with language-specific stopwords.

At indexing time, the language column of each document is examined, and only the stopwords for that language are eliminated. At query time, the session language setting determines the active stopwords, like it determines the active lexer when using the multi-lexer.

**Note:** When indexing a multi-language table with a multi-language stoplist, the table must have a language column.

Examples

**Example 7–2 Single Language Stoplist**

The following example creates a stoplist called mystop:

```sql
begin
    ctx_ddl.create_stoplist('mystop', 'BASIC_STOPLIST');
end;
```

**Example 7–3 Multi-Language Stoplist**

The following example creates a multi-language stoplist called multistop and then adds two language-specific stopwords:

```sql
begin
    ctx_ddl.create_stoplist('multistop', 'MULTI_STOPLIST');
    ctx_ddl.add_stopword('mystop', 'Die', 'german');
    ctx_ddl.add_stopword('mystop', 'Or', 'english');
end;
```
Related Topics

ADD_STOPWORD
ADD_STOPCLASS
ADD_STOPTHEME
DROP_STOPLIST
CREATE_INDEX in Chapter 1, "Oracle Text SQL Statements and Operators".
ALTER_INDEX in Chapter 1, "Oracle Text SQL Statements and Operators".
Appendix E, "Oracle Text Supplied Stoplists"
DROP_INDEX_SET

Drops a CTXCAT index set created with CTX_DDL.CREATE_INDEX_SET.

Syntax

```plsql
CTX_DDL.DROP_INDEX_SET(
    set_name    IN VARCHAR2
);
```

**set_name**
Specify the name of the index set to drop.

Dropping an index set drops all of the sub-indexes it contains.
DROPS a policy created with CTX_DDL.CREATE_POLICY.

Syntax

```
CTX_DDL.DROP_POLICY(
      policy_name  IN VARCHAR2
);
```

**policy_name**
Specify the name of the policy to drop.
The `DROP_PREFERENCE` procedure deletes the specified preference from the Text data dictionary. Dropping a preference does not affect indexes that have already been created using that preference.

**Syntax**

```sql
CTX_DDL.DROP_PREFERENCE(
   preference_name    IN VARCHAR2
);
```

- **preference_name**
  Specify the name of the preference to be dropped.

**Example**

The following example drops the preference `my_lexer`.

```sql
begin
ctx_ddl.drop_preference('my_lexer');
end;
```

**Related Topics**

See also `CTX_DDL.CREATE_PREFERENCE`. 
The DROP_SECTION_GROUP procedure deletes the specified section group, as well as all the sections in the group, from the Text data dictionary.

Syntax

```sql
CTX_DDL.DROP_SECTION_GROUP(
    group_name    IN VARCHAR2
);
```

**group_name**

Specify the name of the section group to delete.

Example

The following example drops the section group `htmgroup` and all its sections:

```sql
begin
    ctx_ddl.drop_section_group('htmgroup');
end;
```

Related Topics

See also CTX_DDL.CREATE_SECTION_GROUP.
DROP_SHADOW_INDEX

Drops a shadow index for the specified index. When you drop a shadow index, if it is partitioned, then its metadata and the metadata of all this shadow index’s partitions are dropped. This procedure also drops all the shadow index tables and cleans up any online pending queue.

Syntax

```sql
CTX_DDL.DROP_SHADOW_INDEX(
    idx_name       in VARCHAR2
);
```

**idx_name**
The name of a valid CONTEXT indextype.

Example

The following example drops the shadow index `myshadowidx`:

```sql
begin
    ctx_ddl.drop_shadow_index('myshadowidx');
end;
```

Related Topics

See also CTX_DDL.CREATE_SHADOW_INDEX.
DROP_STOPLIST

Drops a stoplist from the Text data dictionary. When you drop a stoplist, you must re-create or rebuild the index for the change to take effect.

Syntax

\[
\text{CTX\_DDL\_DROP\_STOPLIST(stoplist\_name \text{ in varchar2});}
\]

**stoplist\_name**
Specify the name of the stoplist.

Example

The following example drops the stoplist *mystop*:

```sql
begin
  ctx_ddl.drop_stoplist('mystop');
end;
```

Related Topics

See also CTX\_DDL\_CREATE\_STOPLIST.
EXCHANGE_SHADOW_INDEX

This procedure swaps the index (or index partition) metadata and index (or index partition) data.

For non-partitioned indexes, this procedure swaps both the metadata and the index data, and processes the online pending queue.

Syntax

```
CTX_DDL.EXCHANGE_SHADOW_INDEX(
    idx_name  IN VARCHAR2
    partition_name  IN VARCHAR2 default NULL
);
```

**idx_name**
Specify the name of the CONTEXT indextype.

**partition_name**
Specify the name of the shadow index partition. May also be NULL.

Example

**Example 7-4  Global Index Recreate with Scheduled Swap**
This example demonstrates running `CTX_DDL.EXCHANGE_SHADOW_INDEX` during non-business hours when query failures and DML blocking can be tolerated.

```sql
/* create lexer and original index */
exec ctx_ddl.create_preference('us_lexer','basic_lexer');
create index idx on tbl(text) indextype is ctxsys.context
    parameters('lexer us_lexer');

/* create a new lexer */
begin
    ctx_ddl.create_preference('e_lexer','basic_lexer');
    ctx_ddl.set_attribute('e_lexer','base_letter','yes');
    ctx_ddl.create_preference('m_lexer','multi_lexer');
    ctx_ddl.add_sub_lexer('m_lexer','default','us_lexer');
    ctx_ddl.add_sub_lexer('m_lexer','e','e_lexer');
end;
/
/* add new language column to the table for multi-lexer */
alter table tbl add(lang varchar2(10) default 'us');

/* recreate index online with the new multip-lexer */
exec ctx_ddl.create_shadow_index('idx',
    'replace lexer m_lexer language column lang');
exec ctx_ddl.exchange_shadow_index('idx');
```

Notes

**Using EXCHANGE_SHADOW_INDEX with Non-partitioned Indexes**
For non-partitioned indexes, this procedure will swap both metadata and index data, and will process the online pending queue.
Queries will return *column not indexed* errors when swapping metadata and index data, but queries are processed normally when processing online pending queue. The period of errors being raised should be short.

If you specify **POPULATE** when you create the shadow index, and if many DML operations have been issued since the creation of the shadow index, then there could be a large pending queue. However, if you use **incremental recreate**, that is, specify **NOPOPULATE** when you create the shadow index, and you then populate the pending queue and sync, then the online pending queue is always empty no matter how many DML operations have occurred since **CREATE_SHADOW_INDEX** was issued.

When this procedure is running, DML will first fail with an error about index being in in-progress status. After that DML could be blocked (hang) if there are rows in online pending queue that need to be reapplied.

---

**Note:** When this procedure is running, DML statements will fail with an error that the index is in "in-progress status." If, when this error occurs, there are rows in the online pending queue that need to be reapplied, then the DML could be blocked and hang.

---

**Using EXCHANGE_SHADOW_INDEX with Partitioned Indexes**

For partitions that are recreated with **NOSWAP**: when the index is partitioned, and if **partition_name** is a valid index partition, then this procedure will swap the index partition data and the index partition metadata, and will process the online pending queue for this partition.

This procedure swaps only one partition at a time. When you run this procedure on partitions that are recreated with **NOSWAP**:

- Queries that span multiple partitions will not return consistent results across all partitions.
- Queries on the partition that is being swapped will return errors.
- Queries on partitions that are already swapped will be based on the new index.
- Queries on the partitions that haven’t been swapped will be based on the old index.

If the **partition_name** is NULL, then this procedure will swap the index metadata. Run this procedure as the last step when recreating a local partitioned index online.

**Related Topics**

See also CTX_DDL.RECREATE_INDEX_ONLINE, CTX_DDL.CREATE_SHADOW_INDEX, and CTX_DDL.DROP_SHADOW_INDEX.
OPTIMIZE_INDEX

Use this procedure to optimize the index. Optimize your index after you synchronize it. Optimizing an index removes old data and minimizes index fragmentation, which can improve query response time. Querying and DML may proceed while optimization takes place.

You can optimize in fast, full, rebuild, token, or token-type mode.

- Fast mode compacts data but does not remove rows.
- Full mode compacts data and removes rows.
- Optimize in rebuild mode rebuilds the $I$ table (the inverted list table) in its entirety. Rebuilding an index is often significantly faster than performing a full optimization, and is more likely to result in smaller indexes, especially if the index is heavily fragmented.

Rebuild optimization creates a more compact copy of the $I$ table, and then switches the original $I$ table and the copy. The rebuild operation will therefore require enough space to store the copy as well as the original. (If redo logging is enabled, then additional space is required in the redo log as well.) At the end of the rebuild operation, the original $I$ table is dropped, and the space can be reused.

Optimize in rebuild mode supports partitioning on the $I$ table via the i_table_clause attribute of the basic_storage preference with the following limitations:

- The i_index_clause must specify using a local btree index if the $I$ table is partitioned.
- Partitioning schemes on the token_first, token_last, or token_count columns are not allowed.

- In token mode, specify a specific token to be optimized (for example, all rows with documents containing the word elections). Use this mode to optimize index tokens that are frequently searched, without spending time on optimizing tokens that are rarely referenced. An optimized token can improve query response time (but only for queries on that token).

- Token-type optimization is similar to token mode, except that the optimization is performed on field sections or MDATA sections (for example, sections with an <A> tag). This is useful in keeping critical field or MDATA sections optimal.

A common strategy for optimizing indexes is to perform regular token optimizations on frequently referenced terms, and to perform rebuild optimizations less frequently. (Use CTX_REPORT.QUERY_LOG_SUMMARY to find out which queries are made most frequently.) You can perform full, fast, or token-type optimizations instead of token optimizations.

Some users choose to perform frequent time-limited full optimizations along with occasional rebuild optimizations.

**Note:** Optimizing an index can result in better response time only if you insert, delete, or update documents in the base table after your initial indexing operation.
Using this procedure to optimize the index is recommended over using the `ALTER INDEX` statement.

Optimization of a large index may take a long time. To monitor the progress of a lengthy optimization, log the optimization with `CTX_OUTPUT.START_LOG` and check the resultant logfile from time to time.

Note that, unlike serial optimize full, `CTX_DDL.OPTIMIZE_INDEX()` run with `optlevel` of `FULL` and `parallel_degree > 1` is not resumable. That is, it will not resume from where it left after a time-out or failure.

---

**Note:** There is a very small window of time when a query might fail in `CTX_DDL.OPTIMIZE_INDEX` REBUILD mode when the $I$ table is being swapped with the optimized shadow $I$ table.

---

**Syntax**

```sql
CTX_DDL.OPTIMIZE_INDEX(
    idx_name        IN  VARCHAR2,
    optlevel        IN  VARCHAR2,
    maxtime         IN  NUMBER DEFAULT NULL,
    token           IN VARCHAR2 DEFAULT NULL,
    part_name       IN VARCHAR2 DEFAULT NULL,
    token_type      IN NUMBER DEFAULT NULL,
    parallel_degree IN NUMBER DEFAULT 1);
```

**idx_name**

Specify the name of the index. If you do not specify an index name, then Oracle Text chooses a single index to optimize.

**optlevel**

Specify optimization level as a string. You can specify one of the following methods for optimization:

<table>
<thead>
<tr>
<th>optlevel value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>FAST</code> or <code>CTX_DDL.OPTLEVEL_FAST</code></td>
<td>This method compacts fragmented rows. However, old data is not removed.</td>
</tr>
<tr>
<td></td>
<td>FAST optimization is not supported for <code>CTXCAT</code> indexes. FAST optimization will not optimize <code>$S$</code> index table.</td>
</tr>
<tr>
<td><code>FULL</code> or <code>CTX_DDL.OPTLEVEL_FULL</code></td>
<td>In this mode you can optimize the entire index or a portion of the index. This method compacts rows and removes old data (deleted rows). Optimizing in full mode runs even when there are no deleted rows.</td>
</tr>
<tr>
<td></td>
<td>Full optimization is not supported for <code>CTXCAT</code> indexes.</td>
</tr>
<tr>
<td><code>REBUILD</code> or <code>CTX_DDL.OPTLEVEL_REBUILD</code></td>
<td>This optlevel rebuilds the $I$ table (the inverted list table) to produce more compact token info rows. Like <code>FULL</code> optimize, this mode also deletes information pertaining to deleted rows of the base table.</td>
</tr>
<tr>
<td></td>
<td><code>REBUILD</code> is not supported for <code>CTXCAT, CTXRULE, or CTXXPATH</code> indexes. <code>REBUILD</code> optimization is also not supported for <code>CONTEXT</code> indexes that have substring indexing enabled.</td>
</tr>
<tr>
<td></td>
<td><code>REBUILD</code> is not supported when the $I$ table is partitioned.</td>
</tr>
</tbody>
</table>
The behavior of CTX_DDL.OPTIMIZE_INDEX with respect to the $S index table is as follows:

<table>
<thead>
<tr>
<th>optlevel value</th>
<th>Will Optimize $S Index Table</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAST or CTX_DDL.OPTLEVEL_FAST</td>
<td>No</td>
<td>■ The optimize process will optimize $I table first. Once $I table optimize is finished, CTX_DDL.OPTIMIZE_INDEX will continue on to optimize $S index table.</td>
</tr>
<tr>
<td>FULL or CTX_DDL.OPTLEVEL_FULL</td>
<td>Yes</td>
<td>■ MAXTIME will also be honored. Once CTX_DDL.OPTIMIZE_INDEX completes optimizing $S rows for a given SDATA_ID, it will check MAXTIME and exit if total elapsed time (including time taken to optimize $I) exceeds specified MAXTIME. The next CTX_DDL.OPTIMIZE_INDEX with optlevel=&gt;'FULL' will pick up where it left off.</td>
</tr>
<tr>
<td>REBUILD or CTX_DDL.OPTLEVEL_REBUILD</td>
<td>Yes</td>
<td>■ $S table optimize will be done in serial.</td>
</tr>
</tbody>
</table>

Note: If for some reason $S optimize exits abnormally, then it is recommended that you use optlevel=>'TOKEN_TYPE' to optimize $S to avoid rebuilding the $I table again.

■ $S table optimize will be done in serial.
maxtime
Specify maximum optimization time, in minutes, for FULL optimize.
When you specify the symbol CTX_DDL.MAXTIME_UNLIMITED (or pass in NULL), the entire index is optimized. This is the default.

token
Specify the token to be optimized.

part_name
If your index is a local index, then you must specify the name of the index partition to synchronize otherwise an error is returned.
If your index is a global, non-partitioned index, then specify NULL, which is the default.

token_type
Specify the token_type to be optimized.

parallel_degree
Specify the parallel degree as a number for parallel optimization. The actual parallel degree depends on your resources. Note that when using REBUILD, setting parallel_degree to a value greater than 1 still results in serial execution.
Because the following optlevel values are executed serially, this setting is ignored for them:

- TOKEN or CTX_DDL.OPTLEVEL_TOKEN
- FAST or CTX_DDL.OPTLEVEL_FAST

Examples
The following two examples are equivalent ways of optimizing an index using fast optimization:

```sql
begin
    ctx_ddl.optimize_index('myidx','FAST');
end;
```

```sql
begin
    ctx_ddl.optimize_index('myidx',CTX_DDL.OPTLEVEL_FAST);
end;
```

The following example optimizes the index token Oracle:

```sql
begin
    ctx_ddl.optimize_index('myidx','token', TOKEN=>'Oracle');
end;
```
To optimize all tokens of field section MYSEC in index MYINDEX:

```sql
begin
  ctx_ddl.optimize_index('myindex', ctx_ddl.optlevel_token_type,
                      token_type=> ctx_report.token_type('myindex','field mysec text'));
end;
```

**Notes**

You can run `CTX_DDL.SYNC` and `CTX_DDL.OPTIMIZE` at the same time. You can also run `CTX_DDL.SYNC` and `CTX_DDL.OPTIMIZE` with parallelism at the same time. However, you should not:

- Run `CTX_DDL.SYNC` with parallelism at the same time as `CTX_DDL.OPTIMIZE` with parallelism.

If you should run one of these combinations, no error is generated; however, one operation will wait until the other is done.

**Related Topics**

See also `CTX_DDL.SYNC_INDEX` and `ALTER INDEX` in Chapter 1, "Oracle Text SQL Statements and Operators".
This procedure populates the pending queue with every rowid in the base table or table partition. This procedure is only supported for CONTEXT indexes.

This procedure is valuable for large installations that cannot afford to have the indexing process run continuously, and, therefore, need finer control over creating text indexes. The preferred method is to create an empty index, place all the rowids into the pending queue, and build the index through CTX_DDL.SYNC_INDEX.

Syntax

```sql
ctx_ddl.populate_pending(
    idx_name  IN VARCHAR2,
    part_name IN VARCHAR2 DEFAULT NULL
);
```

idx_name
Name of the CONTEXT indextype.

part_name
Name of the index partition, if any. Must be provided for local partitioned indexes and must be NULL for global, non-partitioned indexes.

Notes

The SYNC_INDEX is blocked for the duration of the processing. The index unit must be totally empty (idx_docid_count = 0, idx_nextid = 1). The rowids of rows waiting to be indexed are inserted into table ctxsys.dr$pending. You should ensure that there is sufficient space in this table to hold the rowids of the base table.

Related Topics

See also SYNC_INDEX, CREATE_SHADOW_INDEX, DROP_SHADOW_INDEX, EXCHANGE_SHADOW_INDEX, RECREATE_INDEX_ONLINE.
RECREATE_INDEX_ONLINE

Recreates the specified index, or recreates the passed-in index partition if the index is local partitioned. For global non-partitioned indexes, this is a one-step procedure. For local partitioned indexes, this procedure must be run separately on every partition after first using CREATE_SHADOW_INDEX to create a shadow policy (or metadata). This procedure is only supported in Enterprise Edition of Oracle Database.

The following changes are not supported:

- Transitioning from non-composite domain index to composite, or changing the composite domain index columns.
- Rebuilding indexes that have partitioned index tables, for example, $I, $P, $K.

Syntax

```sql
CTX_DDL.RECREATE_INDEX_ONLINE(
    idx_name          IN VARCHAR2,
    parameter_string  IN VARCHAR2 default NULL,
    parallel_degree   IN NUMBER default 1,
    partition_name    IN VARCHAR2 default NULL
);
```

**idx_name**
The name of a valid CONTEXT indextype.

**parameter_string**
If the index is a global non-partitioned index, specify the same index-level parameter string as in ALTER INDEX. Must start with REPLACE, if it is not NULL. Optionally specify SWAP or NOSWAP. The default is SWAP.

**parallel_degree**
Reserved for future use. Specify the degree of parallelism. Parallel operation is not supported in the current release.

**partition_name**
Specify the name of a valid index partition for a local partitioned index. Otherwise, the default is NULL. If the index is partitioned, then first pass a partition name, and then specify the partition-level parameter string for ALTER INDEX REBUILD PARTITION.

Examples

**Example 7–5  Recreate Simple Global Index**
The following example creates an index `idx` with a BASIC_LEXER-based preference `us_lexer`. It then recreates the index with a new MULTI_LEXER based preference `m_lexer` in one step. You can use this one step approach when you do not mind that a query might fail for a small window of time at the end of the operation, and DML might get blocked at the beginning for a short time and again at the end.

```sql
/* create lexer and original index */
exec ctx_ddl.create_preference('us_lexer', 'basic_lexer');
create index idx on tbl(text) indextype is ctxsys.context
    parameters('lexer us_lexer');

/* create a new lexer */
begin
```
ctx_ddl.create_preference('e_lexer','basic_lexer');
cctx_ddl.set_attribute('e_lexer','base_letter','yes');
cctx_ddl.create_preference('m_lexer','multi_lexer');
cctx_ddl.add_sub_lexer('m_lexer','default','us_lexer');
cctx_ddl.add_sub_lexer('m_lexer','e','e_lexer');
end;
/
/* add new language column to the table for multi-lexer */
alter table tbl add(lang varchar2(10) default 'us');
/* recreate index online with the new multi-lexer */
exec ctx_ddl.recreate_index_online('idx',
  'replace lexer m_lexer language column lang');

**Example 7–6  Local Index Recreate with All-At-Once Swap**

The following example creates a local partitioned index `idxp` with a basic lexer `us_lexer`. It has two index partitions `idx_p1` and `idx_p2`. It then recreates a local partitioned index `idxp` online with partition `idx_p1`, which will have a new storage preference `new_store`. The swapping of the partition metadata and index partition data occur at the end. In this example, queries spanning multiple partitions return consistent results across partitions when recreate is in process, except at the end when `EXCHANGE_SHADOW_INDEX` is running. The extra space required is the combined index size of partition `idx_p1` and `idx_p2`.

/* create a basic lexer and a local partition index with the lexer*/
exec ctx_ddl.create_preference('us_lexer','basic_lexer');
ccreate index idxp on tbl text) indextype is ctxsys.context local
  (partition idx_p1,
   partition idx_p2)
parameters('lexer us_lexer');
/* create new preferences */
begin
cctx_ddl.create_preference('my_store','basic_storage');
cctx_ddl.set_attribute('my_store','i_table_clause','tablespace tbs');
end;
/
begin
cctx_ddl.create_preference('e_lexer','basic_lexer');
cctx_ddl.set_attribute('e_lexer','base_letter','yes');
cctx_ddl.create_preference('m_lexer','multi_lexer');
cctx_ddl.add_sub_lexer('m_lexer','default','us_lexer');
cctx_ddl.add_sub_lexer('m_lexer','e','e_lexer');
end;
/
/* add new language column */
alter table tbl add column (lang varchar2(10) default 'us');
/* create a shadow policy with a new lexer */
exec ctx_ddl.create_shadow_index('idxp', null,
  'replace lexer m_lexer language column lang');
/* recreate every index partition online without swapping */
exec ctx_ddl.recreate_index_online('idxp',
  'replace storage my_store NOSWAP', 1, 'idx_p1');
exec ctx_ddl.recreate_index_online('idxp', 'replace NOSWAP',1,'idx_p2');
Example 7–7  Local Index Recreate with Per-Partition Swap
This example performs the same tasks as Example 7–6, “Local Index Recreate with All-At-Once Swap”, except that each index partition is swapped in as it is completed. Queries across all partitions may return inconsistent results in this example.

    /* create a basic lexer and a local partition index with the lexer*/
exec ctx_ddl.create_preference('us_lexer','basic_lexer');
create index idxp on tblp(text) indextype is ctxsys.context local
    (partition idx_p1,
     partition idx_p2)
parameters('lexer us_lexer');
/* create new preferences */
begin
    ctx_ddl.create_preference('my_store','basic_storage');
    ctx_ddl.set_attribute('my_store','i_table_clause','tablespace tbs');
end;
/
begin
    ctx_ddl.create_preference('e_lexer','basic_lexer');
    ctx_ddl.set_attribute('e_lexer','base_letter','yes');
    ctx_ddl.create_preference('m_lexer','multi_lexer');
    ctx_ddl.add_sub_lexer('m_lexer','default','us_lexer');
    ctx_ddl.add_sub_lexer('m_lexer','e','e_lexer');
end;
/
/* add new language column */
alter table tblp add column (lang varchar2(10) default 'us');
/* create a shadow policy with a new lexer */
exec ctx_ddl.create_shadow_index('idxp',
    'replace lexer m_lexer language column lang');
/* recreate every index partition online and swap (default) */
exec ctx_ddl.recreate_index_online('idxp',
    'replace storage my_store', 1, 'idx_p1');
exec ctx_ddl.recreate_index_online('idxp', 'replace SWAP', 1, 'idx_p2',
/* exchange in shadow index metadata */
exec ctx_ddl.exchange_shadow_index('idxp')

Example 7–8  Scheduled Local Index Recreate with All-At-Once Swap
This example shows the incremental recreation of a local partitioned index, where partitions are all swapped at the end.

    /* create a basic lexer and a local partition index with the lexer*/
exec ctx_ddl.create_preference('us_lexer','basic_lexer');
create index idxp on tblp(text) indextype is ctxsys.context local
(partition idx_p1,  
  partition idx_p2)  
parameters('lexer us_lexer');  
  
/* create new preferences */  
begin  
  ctx_ddl.create_preference('my_store','basic_storage');  
  ctx_ddl.set_attribute('my_store','i_table_clause','tablespace tbs');  
end;  
  
/* create a shadow policy with a new lexer */  
exec ctx_ddl.create_shadow_index('idxp', 'replace lexer m_lexer language column lang');  
/* create shadow partition with new storage preference */  
exec ctx_ddl.recreate_index_online('idxp', 'replace storage ctxsys.default_storage nopopulate',1,'idx_p1');  
exec ctx_ddl.recreate_index_online('idxp', 'replace storage ctxsys.default_storage nopopulate',1,'idx_p2');  
  
declare  
  idxid integer;  
  ixpid integer;  
begin  
  select idx_id into idxid from ctx_user_indexes  
    where idx_name = 'IDXP';  
  select ixp_id into ixpid from ctx_user_index_partitions  
    where ixp_index_name = 'IDXP'  
      and ixp_index_partition_name = 'IDX_P1';  
  /* populate pending */  
  ctx_ddl.populate_pending('RIO$'||idxid, 'RIO$'||idxid||'#'||ixpid);  
  /* incremental sync */  
  ctx_ddl.sync_index('RIO$'||idxid, null, 'RIO$'||idxid||'#'||ixpid, maxtime=>400);  
  /* more incremental sync until no more pending rows */  
  select ixp_id into ixpid from ctx_user_index_partitions  
    where ixp_index_name = 'IDXP'  
      and ixp_index_partition_name = 'IDX_P2';  
  /* populate pending */  
  ctx_ddl.populate_pending('RIO$'||ixpid, 'RIO$'||ixpid||'#'||ixpid);  
  /* incremental sync */  
  ctx_ddl.sync_index('RIO$'||ixpid, null, 'RIO$'||ixpid||'#'||ixpid, maxtime=>400);  
  /* more incremental sync until no more pending rows */  
end;  
  
exec ctx_ddl.exchange_shadow_index('idxp','idx_p1');
exec ctx_ddl.exchange_shadow_index('idxp','idx_p2');
exec ctx_ddl.exchange_shadow_index('idxp');

**Example 7–9  Schedule Local Index Recreate with Per-Partition Swap**

For incremental recreate where partitions are swapped as they becomes available, follow the steps in example Example 7–8, "Scheduled Local Index Recreate with All-At-Once Swap", except instead of waiting until all syncs are finished before starting exchange shadow index, EXCHANGE_SHADOW_INDEX is done for each partition right after sync is finished.

**Notes**

**Using RECREATE_INDEX_ONLINE with Global Non-partitioned Indexes**

For global indexes, this procedure provides a one-step process to recreate an index online. It recreates an index, with new preference values, while preserving base table DML and query capability during the recreate process.

---

**Note:** Because the new index is created alongside the existing index, this operation requires additional storage roughly equal to the size of the existing index.

---

**DML Behavior**

Because this procedure is performed online, DML on the base table are permitted during this operation, and are processed as normal. All DML statements that occur during RECREATE_INDEX_ONLINE are logged into an online pending queue.

Towards the end of the recreate operation, there will be a short duration when DML will fail with an error being raised stating that the index is in an in-progress status. DML may hang again during the process, and the duration will depend on how many DML are logged in the online pending queue since the start of the recreate process.

Note that after the recreate index operation is complete, new information, from all the DML that becomes pending since RECREATE_INDEX_ONLINE started, may not be immediately reflected. As with creating an index with INDEXTYPE IS ctxsys.context ONLINE, the index should be synchronized after the recreate index operation is complete, to bring it fully up-to-date.

**See Also:**

- CTX_DDL.CREATE_SHADOW_INDEX and CTX_DDL.EXCHANGE_SHADOW_INDEX for information about how to manually go through each stage of recreation, and to schedule each step to run at a preferred time
- The ONLINE parameter under "Syntax for CONTEXT Index Type" on page 1-37

**Sync and Optimize Behavior**

Syncs issued against the index during the recreate operation are processed against the old, existing data. Syncs are also blocked during the same window when queries return errors. Optimize commands issued against the index during the recreate operation return immediately without error and without processing.
Query Behavior
During the recreate operation, the index can be queried normally most of the time. Queries return results based on the existing index and policy (or metadata) until after the final swap.

There is a short interval towards the end of `RECREATE_INDEX_ONLINE` when queries will return an error indicating that the column is not indexed. This duration should be short for regular queries. It is mainly the time taken for swapping data segments of the shadow index tables and the index tables, plus the time to delete all the rows in the pending queue. This is the same window of time when DML will fail.

During `RECREATE_INDEX_ONLINE`, if you issue DML statements and synchronize them, then you will be able to see the new rows when you query on the existing index. However, after `RECREATE_INDEX_ONLINE` finishes (swapping completes and query is on the new index) and before sync is performed, it is possible that you will not be able to query on the new rows, which once could be queried on the old index.

Note: Transactional queries are not supported.

Using `RECREATE_INDEX_ONLINE` with Local Partitioned Indexes
If the index is local partitioned, you cannot recreate index in one step. You must first create a shadow policy, and then run this procedure for every partition. You can specify SWAP or NOSWAP to indicate whether `RECREATE_INDEX_ONLINE` partition will swap the index partition data and index partition metadata or not. If the partition was built with NOSWAP, then another call to `EXCHANGE_SHADOW_INDEX` must be invoked later against this partition.

This procedure can also be used to update the metadata (for example, storage preference) of each partition when you specify NOPOPULATE in the parameter string. This is useful for incremental building of a shadow index through time-limited sync.

If NOPOPULATE is specified, then NOSWAP is silently enforced.

NOSWAP Behavior
During the recreate of the index partition, since no swapping is performed, queries on the partition are processed regularly. Until the swapping stage is reached, queries spanning multiple partitions return consistent results across partitions.

DML and sync are processed normally. Running optimize on partitions that are being recreated, or that have been built (but not swapped), simply returns without doing anything. Running optimize on a partition that has not been rebuilt processes normally.

As with a global index, when all of the partitions use NOSWAP, the additional storage requirement is roughly equal to the size of the existing index.

SWAP Behavior
Because index partition data and metadata are swapped after index recreate, queries that span multiple partitions will not return consistent results from partition to partition, but will always be correct with respect to each index partition. There is also a short interval towards the end of partition recreate, when the index partition is swapped, during which a query will return a "column not indexed" error.

When partitions are recreated with SWAP, the additional storage requirement for the operation is equal to the size of the existing index partition.

DML on the partition is blocked. Sync is also blocked during swapping.
Related Topics

See also `CREATE_SHADOW_INDEX` and `DROP_SHADOW_INDEX`, and `EXCHANGE_SHADOW_INDEX`, as well as *Oracle Text Application Developer’s Guide*. 
REMOVE_INDEX

Removes the index with the specified column list from a CTXCAT index set preference.

Syntax

```
CTX_DDL.REMOVE_INDEX(
    set_name       IN VARCHAR2,
    column_list    IN VARCHAR2
    language       IN VARCHAR2 default NULL
);
```

**set_name**
Specify the name of the index set.

**column_list**
Specify the name of the column list to remove.

Note: This procedure does not remove a CTXCAT sub-index from the existing index. To do so, you must drop your index and re-index with the modified index set preference.
REMOVE_MDATA

Use this procedure to remove metadata values, which are associated with an MDATA section, from a document. Only the owner of the index is allowed to call ADD_MDATA and REMOVE_MDATA.

Syntax

```sql
CTX_DDL.REMOVE_MDATA(
    idx_name           IN VARCHAR2,
    section_name       IN VARCHAR2,
    values             SYS.ODCIVARCHAR2LIST,
    rowids             SYS.ODCIRIDLIST,
    [part_name]        IN VARCHAR2
);
```

**idx_name**
Name of the text index that contains the named rowids.

**section_name**
Name of the MDATA section.

**values**
List of metadata values. If a metadata value contains a comma, the comma must be escaped with a backslash.

**rowids**
Rowids from which to remove the metadata values.

**[part_name]**
Name of the index partition, if any. Must be provided for local partitioned indexes and must be NULL for global, non-partitioned indexes.

Example

This example removes the MDATA value `blue` from the MDATA section `BGCOLOR`.

```sql
ctx_ddl.remove_mdata('idx_docs', 'bgcolor', 'blue', 'rows');
```

Related Topics

See also "ADD_MDATA" on page 7-9; "ADD_MDATA_SECTION" on page 7-12; "MDATA" on page 3-28; as well as the Section Searching chapter of Oracle Text Application Developer's Guide.

Notes

These updates are updates directly on the index itself, not on the actual contents stored in the base table. Therefore, they will not survive when the Text index is rebuilt.
The **REMOVE_SECTION** procedure removes the specified section from the specified section group. You can specify the section by name or ID. View section ID with the **CTX_USER_SECTIONS** view.

**Syntax 1**

Use the following syntax to remove a section by section name:

```sql
CTX_DDL.REMOVE_SECTION(
    group_name       in    varchar2,
    section_name     in    varchar2
);
```

**group_name**
Specify the name of the section group from which to delete **section_name**.

**section_name**
Specify the name of the section to delete from **group_name**.

**Syntax 2**

Use the following syntax to remove a section by section ID:

```sql
CTX_DDL.REMOVE_SECTION(
    group_name     in    varchar2,
    section_id     in    number
);
```

**group_name**
Specify the name of the section group from which to delete **section_id**.

**section_id**
Specify the section ID of the section to delete from **group_name**.

**Example**

The following example drops a section called **Title** from the **htmgroup**:

```sql
begin
    ctxt_ddl.remove_section('htmgroup', 'Title');
end;
```

**Related Topics**

- **ADD_FIELD_SECTION**
- **ADD_SPECIAL_SECTION**
- **ADD_ZONE_SECTION**
**REMOVE_STOPCLASS**

Removes a stopclass from a stoplist.

**Syntax**

```sql
CTX_DDL.REMOVE_STOPCLASS(
    stoplist_name  in   varchar2,
    stopclass      in   varchar2
);
```

**stoplist_name**
Specify the name of the stoplist.

**stopclass**
Specify the name of the stopclass to be removed.

**Example**

The following example removes the stopclass NUMBERS from the stoplist mystop.

```sql
begin
    ctx_ddl.remove_stopclass('mystop', 'NUMBERS');
end;
```

**Related Topics**

ADD_STOPCLASS
REMOVE_STOPTHEME

Removes a stoptheme from a stoplist.

Syntax

```sql
CTX_DDL.REMOVE_STOPTHEME(
    stoplist_name   in   varchar2,
    stoptheme       in   varchar2
);
```

**stoplist_name**
Specify the name of the stoplist.

**stoptheme**
Specify the stoptheme to be removed from stoplist_name.

Example

The following example removes the stoptheme `banking` from the stoplist `mystop`:

```sql
begin
    ctx_ddl.remove_stoptheme('mystop', 'banking');
end;
```

Related Topics

ADD_STOPTHEME
REMOVE_STOPWORD

Removes a stopword from a stoplist. To have the removal of a stopword be reflected in the index, you must rebuild your index.

Syntax

```sql
CTX_DDL.REMOVE_STOPWORD(
  stoplist_name  in   varchar2,
  stopword       in   varchar2,
  language       in   varchar2 default NULL
);
```

**stoplist_name**

Specify the name of the stoplist.

**stopword**

Specify the stopword to be removed from `stoplist_name`.

**language**

Specify the language of `stopword` to remove when the stoplist you specify with `stoplist_name` is of type `MULTI_STOPLIST`. You must specify the globalization support name or abbreviation of an Oracle Text-supported language. You can also remove ALL stopwords.

Example

The following example removes a stopword `because` from the stoplist `mystop`:

```sql
begin
  ctx_ddl.remove_stopword('mystop','because');
end;
```

Related Topics

ADD_STOPWORD
**REMOVE_SUB_LEXER**

Removes a sub-lexer from a multi-lexer preference. You cannot remove the lexer for DEFAULT.

**Syntax**

```
CTX_DDL.REMOVE_SUB_LEXER(
lexer_name  in   varchar2,
language    in   varchar2 default NULL
);
```

**lexer_name**
Specify the name of the multi-lexer preference.

**language**
Specify the language of the sub-lexer to remove. You must specify the globalization support name or abbreviation of an Oracle Text-supported language.

**Example**

The following example removes a sub-lexer `germanLexer` of language `german`:

```
begin
  ctx_ddl.remove_sub_lexer('germanLexer','german');
end;
```

**Related Topics**

ADD_SUB_LEXER
REPLACE_INDEX_METADATA

Use this procedure to replace metadata in local domain indexes at the global (index) level.

Note: The ALTER_INDEX_PARAMETERS command performs the same function as this procedure and can replace more than just metadata. For that reason, using ALTER_INDEX_PARAMETERS is the preferred method of replacing metadata at the global (index) level and should be used in place of this procedure when possible. For more information, see "ALTER_INDEX_PARAMETERS Syntax" on page 1-3.

CTX_REPLACE_INDEX_METADATA may be deprecated in a future release of Oracle Text.

Syntax

```
CTX_DDL.REPLACE_INDEX_METADATA(idx_name IN VARCHAR2,
     parameter_string IN VARCHAR2);
```

**idx_name**
Specify the name of the index whose metadata you want to replace.

**parameter_string**
Specify the parameter string to be passed to ALTER_INDEX. This must begin with 'REPLACE METADATA'.

Notes

ALTER_INDEX_REBUILD_PARAMETERS ("REPLACE METADATA") does not work for a local partitioned index at the index (global) level; you cannot, for example, use that ALTER_INDEX syntax to change a global preference, such as filter or lexer type, without rebuilding the index. Therefore, CTX_DDL.REPLACE_INDEX_METADATA is provided as a method of overcoming this limitation of ALTER_INDEX.

Though it is meant as a way to replace metadata for a local partitioned index, CTX_DDL.REPLACE_INDEX_METADATA can be used on a global, non-partitioned index, as well.

REPLACE_INDEX_METADATA cannot be used to change the sync type at the partition level; that is, **parameter_string** cannot be 'REPLACE METADATA SYNC'. For that purpose, use ALTER_INDEX_REBUILD_PARTITION to change the sync type at the partition level.

Related Topics

See also "ALTER_INDEX_PARAMETERS Syntax" on page 1-3 and "ALTER_INDEX_REBUILD Syntax" on page 1-4.
Sets a preference attribute. Use this procedure after you have created a preference with CTX_DDL.CREATE_PREFERENCE.

Syntax

```sql
CTX_DDL.SET_ATTRIBUTE(preference_name IN VARCHAR2,
                      attribute_name  IN VARCHAR2,
                      attribute_value IN VARCHAR2);
```

**preference_name**
Specify the name of the preference.

**attribute_name**
Specify the name of the attribute.

**attribute_value**
Specify the attribute value. Specify boolean values as TRUE or FALSE, T or F, YES or NO, Y or N, ON or OFF, or 1 or 0.

Example

**Specifying File Data Storage**

The following example creates a data storage preference called `filepref` that tells the system that the files to be indexed are stored in the operating system. The example then uses CTX_DDL.SET_ATTRIBUTE to set the PATH attribute to the directory `/docs`.

```sql
begin
  ctx_ddl.create_preference('filepref', 'FILE_DATASTORE');
  ctx_ddl.set_attribute('filepref', 'PATH', '/docs');
end;
```

See Also: For more information about data storage, see "Datastore Types" in Chapter 2, "Oracle Text Indexing Elements"

For more examples of using SET_ATTRIBUTE, see "CREATE_PREFERENCE" in this chapter
SYNC_INDEX

Synchronizes the index to process inserts, updates, and deletes to the base table.

---

**Note:** Because CTX_DDL.SYNC_INDEX issues implicit commits, calling CTX_DDL.SYNC_INDEX in a trigger is strongly discouraged. Doing so can result in errors being raised, as both SYNC_INDEX and post-commit $R LOB maintenance try to update the same $R LOB.

---

**Syntax**

```sql
CTX_DDL.SYNC_INDEX(
    idx_name       IN  VARCHAR2 DEFAULT NULL
    memory          IN VARCHAR2 DEFAULT NULL,
    part_name       IN VARCHAR2 DEFAULT NULL,
    parallel_degree IN NUMBER DEFAULT 1
    maxtime         IN NUMBER DEFAULT NULL,
    locking         IN NUMBER DEFAULT LOCK_WAIT
);
```

- **idx_name**
  Specify the name of the index to synchronize.

---

**Note:** When `idx_name` is null, all CONTEXT, CTXRULE, and CTXXPATH indexes that have pending changes are synchronized. You must be connected as ctxsys to perform this operation. Each index or index partition is synchronized in sequence, one after the other. Because of this, the individual syncs are performed with locking set to NOWAIT and maxtime set to 0. Any values that you specify for locking or maxtime on the SYNC_INDEX call are ignored. However, the memory and parallel_degree parameters are passed on to the individual synchronizations.

---

**memory**

Specify the runtime memory to use for synchronization. This value overrides the DEFAULT_INDEX_MEMORY system parameter.

The memory parameter specifies the amount of memory Oracle Text uses for the synchronization operation before flushing the index to disk. Specifying a large amount of memory:

- Improves indexing performance because there is less I/O
- Improves query performance and maintenance because there is less fragmentation

Specifying smaller amounts of memory increases disk I/O and index fragmentation, but might be useful when runtime memory is scarce.

**part_name**

If your index is a local index, then you must specify the name of the index partition to synchronize otherwise an error is returned.

If your index is a global, non-partitioned index, then specify NULL, which is the default.
**parallel_degree**
Specify the degree to run parallel synchronize. A number greater than 1 turns on parallel synchronize. The actual degree of parallelism might be smaller depending on your resources.

**maxtime**
Indicate a suggested time limit on the operation, in minutes. `SYNC_INDEX` will process as many documents in the queue as possible within the time limit. The maxtime value of NULL is equivalent to `CTX_DDL.MAXTIME_UNLIMITED`. This parameter is ignored when `SYNC_INDEX` is invoked without an index name, in which case maxtime value of 0 is used instead. The locking parameter is ignored for automatic syncs (that is, `SYNC ON COMMIT` or `SYNC EVERY`).

The time limit specified is treated as approximate. The actual time taken may be somewhat less than or greater than what you specify. The "time clock" for maxtime does not start until the SYNC lock is acquired.

**locking**
Configure how `SYNC_INDEX` deals with the situation where another sync is already running on the same index or index partition. When locking is ignored because `SYNC_INDEX` is invoked without an index name, then locking value of `LOCK_NOWAIT` is used instead. The locking parameter is ignored for automatic syncs (that is, `SYNC ON COMMIT` or `SYNC EVERY`).

The options for locking are:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>CTX_DDL.LOCK_WAIT</code></td>
<td>If another sync is running, wait until the running sync is complete, then begin sync. (In the event of not being able to get a lock, it will wait forever and ignore the maxtime setting.).</td>
</tr>
<tr>
<td><code>CTX_DDL.LOCK_NOWAIT</code></td>
<td>If another sync is running, immediately returns without error.</td>
</tr>
<tr>
<td><code>CTX_DDL.LOCK_NOWAIT_ERROR</code></td>
<td>If another sync is running, error &quot;DRG-51313: timeout while waiting for DML or optimize lock&quot; is raised.</td>
</tr>
</tbody>
</table>

**Example**

The following example synchronizes the index `myindex` with 2 megabytes of memory:

```sql
begin
ctx_ddl.sync_index('myindex', '2M');
end;
```

The following example synchronizes the `part1` index partition with 2 megabytes of memory:

```sql
begin
ctx_ddl.sync_index('myindex', '2M', 'part1');
end;
```

**Notes**

You can run `CTX_DDL.SYNC` and `CTX_DDL.OPTIMIZE` at the same time. You can also run `CTX_DDL.SYNC` and `CTX_DDL.OPTIMIZE` with parallelism at the same time. However, you should not run `CTX_DDL.SYNC` with parallelism at the same time as `CTX_DDL.OPTIMIZE`, nor `CTX_DDL.SYNC` with parallelism at the same time as `CTX_
DDL.OPTIMIZE with parallelism. If you should run one of these combinations, no error is generated; however, one operation will wait until the other is done.

Related Topics

ALTER INDEX in Chapter 1, "Oracle Text SQL Statements and Operators"
UNSET_ATTRIBUTE

Removes a set attribute from a preference.

Syntax

CTX_DDL.UNSET_ATTRIBUTE(preference_name varchar2,
        attribute_name  varchar2);

preference_name
Specify the name of the preference.

attribute_name
Specify the name of the attribute.

Example

Enabling/Disabling Alternate Spelling
The following example shows how you can enable alternate spelling for German and disable alternate spelling with CTX_DDL.UNSET_ATTRIBUTE:

begin
    ctx_ddl.create_preference('GERMAN_LEX', 'BASIC_LEXER');
    ctx_ddl.set_attribute('GERMAN_LEX', 'ALTERNATE_SPELLING', 'GERMAN');
end;

To disable alternate spelling, use the CTX_DDL.UNSET_ATTRIBUTE procedure as follows:

begin
    ctx_ddl.unset_attribute('GERMAN_LEX', 'ALTERNATE_SPELLING');
end;

Related Topics

SET_ATTRIBUTE on page 7-74
UPDATE_POLICY

Updates a policy created with CREATE_POLICY. Replaces the preferences of the policy. Null arguments are not replaced.

Syntax

```sql
CTX_DDL.UPDATE_POLICY(
    policy_name     IN VARCHAR2,
    filter          IN VARCHAR2 DEFAULT NULL,
    section_group   IN VARCHAR2 DEFAULT NULL,
    lexer           IN VARCHAR2 DEFAULT NULL,
    stoplist        IN VARCHAR2 DEFAULT NULL,
    wordlist        IN VARCHAR2 DEFAULT NULL);
```

**policy_name**  
Specify the name of the policy to update.

**filter**  
Specify the filter preference to use.

**section_group**  
Specify the section group to use.

**lexer**  
Specify the lexer preference to use.

**stoplist**  
Specify the stoplist to use.

**wordlist**  
Specify the wordlist to use.
This chapter describes the CTX_DOC PL/SQL package for requesting document services, such as highlighting extracted text or generating a list of themes for a document.

Many of these procedures exist in two versions: those that make use of indexes, and those that do not. Those that do not are called "policy-based" procedures. They are offered because there are times when you may want to use document services on a single document without creating a Context index in advance. Policy-based procedures enable you to do this.

The policy_* procedures mirror the conventional in-memory document services and are used with policy_name replacing index_name, and document of type VARCHAR2, CLOB, BLOB, or BFILE replacing textkey. Thus, you need not create an index to obtain document services output with these procedures.

For the procedures that generate character offsets and lengths, such as HIGHLIGHT and TOKENS, Oracle Text follows USC-2 codepoint semantics.

The CTX_DOC package includes the following procedures and functions:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FILTER</td>
<td>Generates a plain text or HTML version of a document.</td>
</tr>
<tr>
<td>GIST</td>
<td>Generates a Gist or theme summaries for a document.</td>
</tr>
<tr>
<td>HIGHLIGHT</td>
<td>Generates plain text or HTML highlighting offset information for a document.</td>
</tr>
<tr>
<td>IFILTER</td>
<td>Generates a plain text version of binary data. Can be called from a USER_DATASTORE procedure.</td>
</tr>
<tr>
<td>MARKUP</td>
<td>Generates a plain text or HTML version of a document with query terms highlighted.</td>
</tr>
<tr>
<td>PKENCODE</td>
<td>Encodes a composite textkey string (value) for use in other CTX_DOC procedures.</td>
</tr>
<tr>
<td>POLICY_FILTER</td>
<td>Generates a plain text or HTML version of a document, without requiring an index.</td>
</tr>
<tr>
<td>POLICY_GIST</td>
<td>Generates a Gist or theme summaries for a document, without requiring an index.</td>
</tr>
<tr>
<td>POLICY_HIGHLIGHT</td>
<td>Generates plain text or HTML highlighting offset information for a document, without requiring an index.</td>
</tr>
<tr>
<td>POLICY_MARKUP</td>
<td>Generates a plain text or HTML version of a document with query terms highlighted, without requiring an index.</td>
</tr>
<tr>
<td>Name</td>
<td>Description</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>POLICY_SNIPPET</td>
<td>Generates a concordance for a document, based on query terms, without requiring an index.</td>
</tr>
<tr>
<td>POLICY_THEMES</td>
<td>Generates a list of themes for a document, without requiring an index.</td>
</tr>
<tr>
<td>POLICY_TOKENS</td>
<td>Generates all index tokens for a document, without requiring an index.</td>
</tr>
<tr>
<td>SET_KEY_TYPE</td>
<td>Sets CTX_DOC procedures to accept rowid or primary key document identifiers.</td>
</tr>
<tr>
<td>SNIPPET</td>
<td>Generates a concordance for a document, based on query terms, without requiring an index.</td>
</tr>
<tr>
<td>THEMES</td>
<td>Generates a list of themes for a document.</td>
</tr>
<tr>
<td>TOKENS</td>
<td>Generates all index tokens for a document.</td>
</tr>
</tbody>
</table>
FILTER

Use the CTX_DOC.FILTER procedure to generate either a plain text or HTML version of a document. You can store the rendered document in either a result table or in memory. This procedure is generally called after a query, from which you identify the document to be filtered.

**Note:** The resultant HTML document does not include graphics.

### Syntax 1: In-memory Result Storage

```sql
exec CTX_DOC.FILTER(
    index_name  IN VARCHAR2,
    textkey     IN VARCHAR2,
    restab      IN OUT NOCOPY CLOB,
    plaintext   IN BOOLEAN DEFAULT FALSE);
exec CTX_DOC.HIGHLIGHT_CLOB_QUERY(
    index_name IN VARCHAR2,
    textkey IN VARCHAR2,
    text_query IN CLOB,
    restab IN OUT NOCOPY HIGHLIGHT_TAB,
    plaintext IN BOOLEAN DEFAULT FALSE);
```

### Syntax 2: Result Table Storage

```sql
exec CTX_DOC.FILTER(
    index_name  IN VARCHAR2,
    textkey     IN VARCHAR2,
    restab      IN VARCHAR2,
    query_id    IN NUMBER DEFAULT 0,
    plaintext   IN BOOLEAN DEFAULT FALSE);
exec CTX_DOC.HIGHLIGHT_CLOB_QUERY(
    index_name IN VARCHAR2,
    textkey IN VARCHAR2,
    text_query IN CLOB,
    restab IN VARCHAR2,
    query_id IN NUMBER DEFAULT 0,
    plaintext IN BOOLEAN DEFAULT FALSE);
```

**index_name**

Specify the name of the index associated with the text column containing the document identified by `textkey`.

**textkey**

Specify the unique identifier (usually the primary key) for the document.

The `textkey` parameter can be as follows:

- a single column primary key value
- encoded specification for a composite (multiple column) primary key. Use CTX_DOC.PKENCODE
- the rowid of the row containing the document
Toggle between primary key and rowid identification using CTX_DOC.SET_KEY_TYPE.

restab
You can specify that this procedure store the marked-up text to either a table or to an in-memory CLOB.

To store results to a table, specify the name of the table. The table to which you want to store results must exist before you make this call.

See Also: "Filter Table" in Appendix A, "Oracle Text Result Tables" for more information about the structure of the filter result table

To store results in memory, specify the name of the CLOB locator. If restab is NULL, then a temporary CLOB is allocated and returned. You must de-allocate the locator after using it with DBMS_LOB.FREETEMPORARY().

If restab is not NULL, then the CLOB is truncated before the operation.

query_id
Specify an identifier to use to identify the row inserted into restab.

When query_id is not specified or set to NULL, it defaults to 0. You must manually truncate the table specified in restab.

plaintext
Specify TRUE to generate a plaintext version of the document. Specify FALSE to generate an HTML version of the document if you are using the AUTO_FILTER filter or indexing HTML documents.

Example

In-Memory Filter
The following code shows how to filter a document to HTML in memory.

```
declare
mklob clob;
amt number := 40;
line varchar2(80);
begin
ctx_doc.filter('myindex','1', mklob, FALSE);
-- mklob is NULL when passed-in, so ctx-doc.filter will allocate a temporary
-- CLOB for us and place the results there.
dbms_lob.read(mklob, amt, 1, line);
dbms_output.put_line('FIRST 40 CHARS ARE:'||line);
-- have to de-allocate the temp lob
dbms_lob.freetemporary(mklob);
end;
```

Create the filter result table to store the filtered document as follows:

```
create table filtertab (query_id number,
document clob);
```

To obtain a plaintext version of document with textkey 20, enter the following statement:

```
begin
```
ctx_doc.filter('newsindex', '20', 'filtertab', '0', TRUE);
end;
GIST

Use the CTX_DOC.GIST procedure to generate gist and theme summaries for a document. You can generate paragraph-level or sentence-level gists or theme summaries.

**Note:** CTX_DOC.GIST requires an installed knowledge base. A knowledge base may or may not have been installed with Oracle Text. For more information on knowledge bases, see the Oracle Text Application Developer's Guide.

**Syntax 1: In-Memory Storage**

```sql
CTX_DOC.GIST(
    index_name       IN VARCHAR2,
    textkey          IN VARCHAR2,
    restab           IN OUT CLOB,
    glevel           IN VARCHAR2 DEFAULT 'P',
    pov              IN VARCHAR2 DEFAULT 'GENERIC',
    numParagraphs    IN NUMBER DEFAULT 16,
    maxPercent       IN NUMBER DEFAULT 10,
    num_themes       IN NUMBER DEFAULT 50);
```

**Index name**
Specify the name of the index associated with the text column containing the document identified by textkey.

**Textkey**
Specify the unique identifier (usually the primary key) for the document.

The textkey parameter can as follows:

- a single column primary key value
- an encoded specification for a composite (multiple column) primary key. To encode a composite textkey, use the CTX_DOC.PK_ENCODE procedure
- the rowid of the row containing the document

Toggle between primary key and rowid identification using CTX_DOC.SET_KEY_TYPE.

**Restab**
Specify that this procedure store the gist and theme summaries to either a table or to an in-memory CLOB.

**Syntax 2: Result Table Storage**

```sql
CTX_DOC.GIST(
    index_name       IN VARCHAR2,
    textkey          IN VARCHAR2,
    restab           IN VARCHAR2,
    query_id         IN NUMBER DEFAULT 0,
    glevel           IN VARCHAR2 DEFAULT 'P',
    pov              IN VARCHAR2 DEFAULT NULL,
    numParagraphs    IN NUMBER DEFAULT 16,
    maxPercent       IN NUMBER DEFAULT 10,
    num_themes       IN NUMBER DEFAULT 50);
```

**Index name**
Specify the name of the index associated with the text column containing the document identified by textkey.

**Textkey**
Specify the unique identifier (usually the primary key) for the document.
To store results to a table specify the name of an existing table.

**See Also:** "Gist Table" in Appendix A, "Oracle Text Result Tables"

To store results in memory, specify the name of the CLOB locator. If `restab` is NULL, then a temporary CLOB is allocated and returned. You must de-allocate the locator after using it.

If `restab` is not NULL, then the CLOB is truncated before the operation.

**query_id**
Specify an identifier to use to identify the row(s) inserted into `restab`.

**glevel**
Specify the type of gist or theme summary to produce. The possible values are:

- `P` for paragraph
- `S` for sentence

The default is `P`.

**pov**
Specify whether a gist or a single theme summary is generated. The type of gist or theme summary generated (sentence-level or paragraph-level) depends on the value specified for `glevel`.

To generate a gist for the entire document, specify a value of 'GENERIC' for `pov`. To generate a theme summary for a single theme in a document, specify the theme as the value for `pov`.

When using result table storage, if you do not specify a value for `pov`, then this procedure returns the generic gist plus up to 50 theme summaries for the document.

When using in-memory result storage to a CLOB, you must specify a `pov`. However, if you do not specify a `pov`, then this procedure generates only a generic gist for the document.

---

**Note:** The `pov` parameter is case sensitive. To return a gist for a document, specify 'GENERIC' in all uppercase. To return a theme summary, specify the theme exactly as it is generated for the document.

Only the themes generated by THEMES for a document can be used as input for `pov`.

---

**numParagraphs**
Specify the maximum number of document paragraphs (or sentences) selected for the document gist or theme summaries. The default is 16.

---

**Note:** The `numParagraphs` parameter is used only when this parameter yields a smaller gist or theme summary size than the gist or theme summary size yielded by the `maxPercent` parameter.

This means that the system always returns the smallest size gist or theme summary.
**maxPercent**
Specify the maximum number of document paragraphs (or sentences) selected for the document gist or theme summaries as a percentage of the total paragraphs (or sentences) in the document. The default is 10.

---

**Note:** The `maxPercent` parameter is used only when this parameter yields a smaller gist or theme summary size than the gist or theme summary size yielded by the `numParagraphs` parameter.

This means that the system always returns the smallest size gist or theme summary.

---

**num_themes**
Specify the number of theme summaries to produce when you do not specify a value for `pov`. For example, if you specify 10, this procedure returns the top 10 theme summaries. The default is 50.

If you specify 0 or NULL, then this procedure returns all themes in a document. If the document contains more than 50 themes, only the top 50 themes show conceptual hierarchy.

### Examples

**In-Memory Gist**
The following example generates a non-default size generic gist of at most 10 paragraphs. The result is stored in memory in a CLOB locator. The code then de-allocates the returned CLOB locator after using it.

```sql
set serveroutput on;
declare
glob clob;
amt number := 40;
line varchar2(80);
beginn
tx_doc.gist('newsindex','34',glob, pov => 'GENERIC',numParagraphs => 10);
-- glob is NULL when passed-in, so ctx-doc.gist will allocate a temporary
-- CLOB for us and place the results there.
dbms_lob.read(glob, amt, 1, line);
dbms_output.put_line('FIRST 40 CHAR ARE:'||line);
-- have to de-allocate the temp lob
dbms_lob.freetemporary(glob);
end;
```

**Result Table Gists**
The following example creates a gist table called `CTX_GIST`:

```sql
create table CTX_GIST (query_id number,
pov varchar2(80),
gist CLOB);
```

**Gists and Theme Summaries**
The following example returns a default sized paragraph-level gist for document 34 as well as the top 10 theme summaries in the document:
begin
  ctx_doc.gist('newsindex','34','CTX_GIST', 1, num_themes=>10);
end;

The following example generates a non-default size gist of at most 10 paragraphs:
begin
  ctx_doc.gist('newsindex','34','CTX_GIST',1,pov =>'GENERIC',numParagraphs=>10);
end;

The following example generates a gist whose number of paragraphs is at most 10 percent of the total paragraphs in document:
begin
  ctx_doc.gist('newsindex','34','CTX_GIST',1,pov => 'GENERIC',  maxPercent => 10);
end;

**Theme Summary**
The following example returns a paragraph-level theme summary for *insects* for document 34. The default theme summary size is returned.
begin
  ctx_doc.gist('newsindex','34','CTX_GIST',1, pov => 'insects');
end;
Use the `CTX_DOC.HIGHLIGHT` procedure to generate highlight offsets for a document. The offset information is generated for the terms in the document that satisfy the query you specify. These highlighted terms are either the words that satisfy a word query or the themes that satisfy an `ABOUT` query.

You can generate highlight offsets for either plaintext or HTML versions of the document. The table returned by `CTX_DOC.HIGHLIGHT` does not include any graphics found in the original document. Apply the offset information to the same documents filtered with `CTX_DOC.FILTER`.

You usually call this procedure after a query, from which you identify the document to be processed.

You can store the highlight offsets to either an in-memory PL/SQL table or a result table.

See `CTX_DOC.POLICY_HIGHLIGHT` on page 8-24 for a version of this procedure that does not require an index.

### Syntax 1: In-Memory Result Storage

```sql
exec CTX_DOC.HIGHLIGHT(
    index_name  IN VARCHAR2,
    textkey     IN VARCHAR2,
    text_query  IN VARCHAR2,
    restab      IN OUT NOCOPY HIGHLIGHT_TAB,
    plaintext   IN BOOLEAN DEFAULT FALSE);
exec CTX_DOC.HIGHLIGHT_CLOB_QUERY(
    index_name  IN VARCHAR2,
    textkey     IN VARCHAR2,
    text_query  IN CLOB,
    restab      IN OUT NOCOPY HIGHLIGHT_TAB,
    plaintext   IN BOOLEAN DEFAULT FALSE);
```

### Syntax 2: Result Table Storage

```sql
exec CTX_DOC.HIGHLIGHT(
    index_name  IN VARCHAR2,
    textkey     IN VARCHAR2,
    text_query  IN VARCHAR2,
    restab      IN VARCHAR2,
    query_id    IN NUMBER   DEFAULT 0,
    plaintext   IN BOOLEAN DEFAULT FALSE);
exec CTX_DOC.HIGHLIGHT_CLOB_QUERY(
    index_name  IN VARCHAR2,
    textkey     IN VARCHAR2,
    text_query  IN CLOB,
    restab      IN VARCHAR2,
    query_id    IN NUMBER DEFAULT 0,
    plaintext   IN BOOLEAN DEFAULT FALSE);
```

**index_name**

Specify the name of the index associated with the text column containing the document identified by `textkey`. 
Specify the unique identifier (usually the primary key) for the document. The textkey parameter can be as follows:

- a single column primary key value
- encoded specification for a composite (multiple column) primary key. Use the CTX_DOC.PKENCODE procedure.
- the rowid of the row containing the document

Toggle between primary key and rowid identification using CTX_DOC.SET_KEY_TYPE.

Specify the original query expression used to retrieve the document. If NULL, no highlights are generated.

If text_query includes wildcards, stemming, fuzzy matching which result in stopwords being returned, HIGHLIGHT does not highlight the stopwords.

If text_query contains the threshold operator, the operator is ignored. The HIGHLIGHT procedure always returns highlight information for the entire result set.

You can specify that this procedure store highlight offsets to either a table or to an in-memory PL/SQL table.

To store results to a table specify the name of the table. The table must exist before you call this procedure.

See Also: "Highlight Table" in Appendix A, "Oracle Text Result Tables" for more information about the structure of the highlight result table.

To store results to an in-memory table, specify the name of the in-memory table of type CTX_DOC.HIGHLIGHT_TAB. The HIGHLIGHT_TAB datatype is defined as follows:

```sql
type highlight_rec is record {
  offset number,
  length number
};

type highlight_tab is table of highlight_rec index by binary_integer;
```

CTX_DOC.HIGHLIGHT clears HIGHLIGHT_TAB before the operation.

Specify the identifier used to identify the row inserted into restab.

When query_id is not specified or set to NULL, it defaults to 0. You must manually truncate the table specified in restab.

Specify TRUE to generate a plaintext offsets of the document.

Specify FALSE to generate HTML offsets of the document if you are using the AUTO_FILTER filter or indexing HTML documents.
Examples

Create Highlight Table
Create the highlight table to store the highlight offset information:

```sql
create table hightab(query_id number,
                      offset number,
                      length number);
```

Word Highlight Offsets
To obtain HTML highlight offset information for document 20 for the word *dog*:

```sql
begin
  ctx_doc.highlight('newsindex', '20', 'dog', 'hightab', 0, FALSE);
end;
```

Theme Highlight Offsets
Assuming the index `newsindex` has a theme component, obtain HTML highlight offset information for the theme query of *politics* by issuing the following query:

```sql
begin
  ctx_doc.highlight('newsindex', '20', 'about(politics)', 'hightab', 0, FALSE);
end;
```

The output for this statement are the offsets to highlighted words and phrases that represent the theme of *politics* in the document.

Restrictions

`CTX_DOC.HIGHLIGHT` does not support the use of query templates or highlighting XML attribute values.

Related Topics

See Also: "POLICY_HIGHLIGHT" on page 8-24, "MARKUP" on page 8-14, and "SNIPPET" on page 8-36
IFILTER

Use this procedure to filter binary data to text.

This procedure takes binary data (BLOB IN), filters the data with the AUTO_FILTER filter, and writes the text version to a CLOB. (Any graphics in the original document are ignored.) CTX_DOC.IFILTER employs the safe callout, and it does not require an index, as CTX_DOC.FILTER does.

---

**Note:** This procedure will not be supported in future releases. Applications should use CTX_DOC.POLICY_FILTER instead.

---

Requirements

Because CTX_DOC.IFILTER employs the safe callout mechanism, the SQL*Net listener must be running and configured for extproc agent startup.

Syntax

```
CTX_DOC.IFILTER(data IN BLOB, text IN OUT NOCOPY CLOB);
```

- **data**
  Specify the binary data to be filtered.

- **text**
  Specify the destination CLOB. The filtered data is placed in here. This parameter must be a valid CLOB locator that is writable. Passing NULL or a non-writable CLOB will result in an error. Filtered text will be appended to the end of existing content, if any.

Example

The document text used in a MATCHES query can be VARCHAR2 or CLOB. It does not accept BLOB input, so you cannot match filtered documents directly. Instead, you must filter the binary content to CLOB using the AUTO_FILTER filter. Assuming the document data is in bind variable :doc_blob:

```
declare
doc_text clob;
begint
  -- create a temporary CLOB to hold the document text
doc_text := dbms_lob.createtemporary(doc_text, TRUE, DBMS_LOB.SESSION);

  -- call ctx_doc.ifilter to filter the BLOB to CLOB data
txt_doc.ifilter(:doc_blob, doc_text);

  -- now do the matches query using the CLOB version
for c1 in (select * from queries where matches(query_string, doc_text)>0)
  loop
    -- do what you need to do here
  end loop;

  dbms_lob.freetemporary(doc_text);
end;
```

---

**Note:** This procedure will not be supported in future releases. Applications should use CTX_DOC.POLICY_FILTER instead.
The **CTX_DOC.MARKUP** procedure takes a query specification and a document textkey and returns a version of the document in which the query terms are marked up. These marked-up terms are either the words that satisfy a word query or the themes that satisfy an *ABOUT* query.

You can set the marked-up output to be either plaintext or HTML. The marked-up document returned by **CTX_DOC.MARKUP** does not include any graphics found in the original document.

You can use one of the pre-defined tag sets for marking highlighted terms, including a tag sequence that enables HTML navigation.

You usually call **CTX_DOC.MARKUP** after a query, from which you identify the document to be processed.

You can store the marked-up document either in memory or in a result table.

See **CTX_DOC.POLICY_MARKUP** on page 8-26 for a version of this procedure that does not require an index.

Note: Oracle Text does not guarantee well-formed output from **CTX.DOC.MARKUP**, especially for terms that are already marked up with HTML or XML. In particular, unexpected nesting of markup tags may occasionally result.

### Syntax 1: In-Memory Result Storage

```sql
exec CTX_DOC.MARKUP(
  index_name     IN VARCHAR2,
  textkey        IN VARCHAR2,
  text_query     IN VARCHAR2,
  restab         IN OUT NOCOPY CLOB,
  plaintext      IN BOOLEAN   DEFAULT FALSE,
  tagset         IN VARCHAR2  DEFAULT 'TEXT_DEFAULT',
  starttag       IN VARCHAR2  DEFAULT NULL,
  endtag         IN VARCHAR2  DEFAULT NULL,
  prevtag        IN VARCHAR2  DEFAULT NULL,
  nexttag        IN VARCHAR2  DEFAULT NULL);
exec CTX_DOC.MARKUP_CLOB_QUERY(
  index_name     IN VARCHAR2,
  textkey        IN VARCHAR2,
  text_query     IN CLOB,
  restab         IN OUT NOCOPY CLOB,
  plaintext      IN BOOLEAN DEFAULT FALSE,
  tagset         IN VARCHAR2 DEFAULT 'TEXT_DEFAULT',
  starttag       IN VARCHAR2 DEFAULT NULL,
  endtag         IN VARCHAR2 DEFAULT NULL,
  prevtag        IN VARCHAR2 DEFAULT NULL,
  nexttag        IN VARCHAR2 DEFAULT NULL);
```

### Syntax 2: Result Table Storage

```sql
exec CTX_DOC.MARKUP(
  index_name     IN VARCHAR2,
```

---

**MARKUP**

The **CTX_DOC.MARKUP** procedure takes a query specification and a document textkey and returns a version of the document in which the query terms are marked up. These marked-up terms are either the words that satisfy a word query or the themes that satisfy an *ABOUT* query.

You can set the marked-up output to be either plaintext or HTML. The marked-up document returned by **CTX_DOC.MARKUP** does not include any graphics found in the original document.

You can use one of the pre-defined tag sets for marking highlighted terms, including a tag sequence that enables HTML navigation.

You usually call **CTX_DOC.MARKUP** after a query, from which you identify the document to be processed.

You can store the marked-up document either in memory or in a result table.

See **CTX_DOC.POLICY_MARKUP** on page 8-26 for a version of this procedure that does not require an index.

Note: Oracle Text does not guarantee well-formed output from **CTX.DOC.MARKUP**, especially for terms that are already marked up with HTML or XML. In particular, unexpected nesting of markup tags may occasionally result.

### Syntax 1: In-Memory Result Storage

```sql
exec CTX_DOC.MARKUP(
  index_name     IN VARCHAR2,
  textkey        IN VARCHAR2,
  text_query     IN VARCHAR2,
  restab         IN OUT NOCOPY CLOB,
  plaintext      IN BOOLEAN   DEFAULT FALSE,
  tagset         IN VARCHAR2  DEFAULT 'TEXT_DEFAULT',
  starttag       IN VARCHAR2  DEFAULT NULL,
  endtag         IN VARCHAR2  DEFAULT NULL,
  prevtag        IN VARCHAR2  DEFAULT NULL,
  nexttag        IN VARCHAR2  DEFAULT NULL);
exec CTX_DOC.MARKUP_CLOB_QUERY(
  index_name     IN VARCHAR2,
  textkey        IN VARCHAR2,
  text_query     IN CLOB,
  restab         IN OUT NOCOPY CLOB,
  plaintext      IN BOOLEAN DEFAULT FALSE,
  tagset         IN VARCHAR2 DEFAULT 'TEXT_DEFAULT',
  starttag       IN VARCHAR2 DEFAULT NULL,
  endtag         IN VARCHAR2 DEFAULT NULL,
  prevtag        IN VARCHAR2 DEFAULT NULL,
  nexttag        IN VARCHAR2 DEFAULT NULL);
```

### Syntax 2: Result Table Storage

```sql
exec CTX_DOC.MARKUP(
  index_name     IN VARCHAR2,
```
index_name
Specify the name of the index associated with the text column containing the document identified by textkey.

textkey
Specify the unique identifier (usually the primary key) for the document.

The textkey parameter can be as follows:

- A single column primary key value
- Encoded specification for a composite (multiple column) primary key. Use the CTX_DOC.PKENCODE procedure.
- The rowid of the row containing the document

Toggle between primary key and rowid identification using CTX_DOC.SET_KEY_TYPE.

text_query
Specify the original query expression used to retrieve the document.

If text_query includes wildcards, stemming, fuzzy matching which result in stopwords being returned, MARKUP does not highlight the stopwords.

If text_query contains the threshold operator, the operator is ignored. The MARKUP procedure always returns highlight information for the entire result set.

restab
You can specify that this procedure store the marked-up text to either a table or to an in-memory CLOB.

To store results to a table specify the name of the table. The result table must exist before you call this procedure.

See Also: For more information about the structure of the markup result table, see “Markup Table” in Appendix A, “Oracle Text Result Tables”.

exec CTX_DOC.MARKUP_CLOB_QUERY(
    index_name   IN VARCHAR2,
    textkey      IN CLOB,
    text_query   IN VARCHAR2,
    restab       IN VARCHAR2,
    query_id     IN NUMBER DEFAULT 0,
    plaintext    IN BOOLEAN DEFAULT FALSE,
    tagset       IN VARCHAR2 DEFAULT 'TEXT_DEFAULT',
    starttag     IN VARCHAR2 DEFAULT NULL,
    endtag       IN VARCHAR2 DEFAULT NULL,
    prevtag      IN VARCHAR2 DEFAULT NULL,
    nexttag      IN VARCHAR2 DEFAULT NULL);
To store results in memory, specify the name of the CLOB locator. If restab is NULL, a temporary CLOB is allocated and returned. You must de-allocate the locator after using it.

If restab is not NULL, the CLOB is truncated before the operation.

**query_id**
Specify the identifier used to identify the row inserted into restab.

When query_id is not specified or set to NULL, it defaults to 0. You must manually truncate the table specified in restab.

**plaintext**
Specify TRUE to generate plaintext marked-up document. Specify FALSE to generate a marked-up HTML version of document if you are using the AUTO_FILTER filter or indexing HTML documents.

**tagset**
Specify one of the following pre-defined tag sets. The second and third columns show how the four different tags are defined for each tagset:

<table>
<thead>
<tr>
<th>Tagset</th>
<th>Tag</th>
<th>Tag Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEXT_DEFAULT</td>
<td>starttag</td>
<td>&lt;&lt;&lt;</td>
</tr>
<tr>
<td></td>
<td>endtag</td>
<td>&gt;&gt;&gt;</td>
</tr>
<tr>
<td></td>
<td>prevtag</td>
<td></td>
</tr>
<tr>
<td></td>
<td>nexttag</td>
<td></td>
</tr>
<tr>
<td>HTML_DEFAULT</td>
<td>starttag</td>
<td>&lt;B&gt;</td>
</tr>
<tr>
<td></td>
<td>endtag</td>
<td>&lt;/B&gt;</td>
</tr>
<tr>
<td></td>
<td>prevtag</td>
<td></td>
</tr>
<tr>
<td></td>
<td>nexttag</td>
<td></td>
</tr>
<tr>
<td>HTML_NAVIGATE</td>
<td>starttag</td>
<td>&lt;A NAME=ctx%CURNUM&gt;&lt;B&gt;</td>
</tr>
<tr>
<td></td>
<td>endtag</td>
<td>&lt;/B&gt;&lt;/A&gt;</td>
</tr>
<tr>
<td></td>
<td>prevtag</td>
<td>&lt;A HREF=#ctx%PREVNUM&gt;&lt;&lt;/A&gt;</td>
</tr>
<tr>
<td></td>
<td>nexttag</td>
<td>&lt;A HREF=#ctx%NEXTNUM&gt;&gt;&lt;/A&gt;</td>
</tr>
</tbody>
</table>

**starttag**
Specify the character(s) inserted by MARKUP to indicate the start of a highlighted term.

The sequence of starttag, endtag, prevtag and nexttag with respect to the highlighted word is as follows:

... prevtag starttag word endtag nexttag...

**endtag**
Specify the character(s) inserted by MARKUP to indicate the end of a highlighted term.

**prevtag**
Specify the markup sequence that defines the tag that navigates the user to the previous highlight.

In the markup sequences prevtag and nexttag, you can specify the following offset variables which are set dynamically:
### Offset Variable

<table>
<thead>
<tr>
<th>Offset Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>%CURNUM</td>
<td>the current offset number</td>
</tr>
<tr>
<td>%PREVNUM</td>
<td>the previous offset number</td>
</tr>
<tr>
<td>%NEXTNUM</td>
<td>the next offset number</td>
</tr>
</tbody>
</table>

See the description of the HTML_NAVIGATE tag set for an example.

**nexttag**

Specify the markup sequence that defines the tag that navigates the user to the next highlight tag.

Within the markup sequence, you can use the same offset variables you use for prevtag. See the explanation for prevtag and the HTML_NAVIGATE tag set for an example.

## Examples

### In-Memory Markup

The following code takes document *(the dog chases the cat)*, performs the assigned markup on it, and stores the result in memory.

```sql
set serveroutput on

drop table mark_tab;
create table mark_tab (id number primary key, text varchar2(80) );
insert into mark_tab values ('1', 'The dog chases the cat.');
create index mark_tab_idx on mark_tab(text)
    indextype is ctxsys.context parameters ('filter ctxsys.null_filter');

declare
    mklob clob;
    amt number := 40;
    line varchar2(80);
begin
    ctx_doc.markup('mark_tab_idx','1','dog AND cat', mklob);
    -- mklob is NULL when passed-in, so ctx_doc.markup will
    -- allocate a temporary CLOB for us and place the results there.
    dbms_lob.read(mklob, amt, 1, line);
    dbms_output.put_line('FIRST 40 CHARS ARE:'||line);
    -- have to de-allocate the temp lob
    dbms_lob.freetemporary(mklob);
end;
/
```

The output from this example shows what the marked-up document looks like:

**FIRST 40 CHARS ARE:** The <<<dog>>> chases the <<<cat>>>.

### Markup Table

Create the highlight markup table to store the marked-up document as follows:

```sql
create table markuptab (query_id  number,
    document  clob);
```
**Word Highlighting in HTML**

You can also store your MARKUP results in a table. To create HTML highlight markup for the words `dog` or `cat` for document 23, enter the following statement:

```
begin
  ctx_doc.markup(index_name => 'my_index',
                  textkey => '23',
                  text_query => 'dog|cat',
                  restab => 'markuptab',
                  query_id => '1',
                  tagset => 'HTML_DEFAULT');
end;
```

**Theme Highlighting in HTML**

To create HTML highlight markup for the theme of `politics` for document 23, enter the following statement:

```
begin
  ctx_doc.markup(index_name => 'my_index',
                  textkey => '23',
                  text_query => 'about(politics)',
                  restab => 'markuptab',
                  query_id => '1',
                  tagset => 'HTML_DEFAULT');
end;
```

**Restrictions**

`CTXDOC.MARKUP` does not support the use of query templates.

**Related Topics**

See Also: "POLICY_MARKUP" on page 8-26, "HIGHLIGHT" on page 8-10 and "SNIPPET" on page 8-36
PKENCODE

The CTX_DOC.PKENCODE function converts a composite textkey list into a single string and returns the string.

The string created by PKENCODE can be used as the primary key parameter textkey in other CTX_DOC procedures, such as CTX_DOC.THEMES and CTX_DOC.GIST.

Syntax

```sql
CTX_DOC.PKENCODE(
    pk1    IN VARCHAR2,
    pk2    IN VARCHAR2 DEFAULT NULL,
    pk4    IN VARCHAR2 DEFAULT NULL,
    pk5    IN VARCHAR2 DEFAULT NULL,
    pk6    IN VARCHAR2 DEFAULT NULL,
    pk7    IN VARCHAR2 DEFAULT NULL,
    pk8    IN VARCHAR2 DEFAULT NULL,
    pk9    IN VARCHAR2 DEFAULT NULL,
    pk10   IN VARCHAR2 DEFAULT NULL,
    pk11   IN VARCHAR2 DEFAULT NULL,
    pk12   IN VARCHAR2 DEFAULT NULL,
    pk13   IN VARCHAR2 DEFAULT NULL,
    pk14   IN VARCHAR2 DEFAULT NULL,
    pk15   IN VARCHAR2 DEFAULT NULL,
    pk16   IN VARCHAR2 DEFAULT NULL)
RETURN VARCHAR2;
```

Each PK argument specifies a column element in the composite textkey list. You can encode at most 16 column elements.

Returns

String that represents the encoded value of the composite textkey.

Example

```sql
begin
    ctx_doc.gist('newsindex',CTX_DOC.PKENCODE('smith', 14), 'CTX_GIST');
end;
```

In this example, smith and 14 constitute the composite textkey value for the document.
POLICY_FILTER

Generates a plain text or an HTML version of a document. With this procedure, no CONTEXT index is required.

This procedure uses a trusted callout.

Syntax

```sql
ctx_doc.policy_filter(policy_name in VARCHAR2,
document       in [VARCHAR2|CLOB|BLOB|BFILE],
restab         in out nocopy CLOB,
plaintext      in BOOLEAN default FALSE,
language       in VARCHAR2 default NULL,
format         in VARCHAR2 default NULL,
charset        in VARCHAR2 default NULL);
```

**policy_name**
Specify the policy name created with CTX_DDL.CREATE_POLICY.

**document**
Specify the document to filter.

**restab**
Specify the name of the CLOB locator.

**plaintext**
Specify TRUE to generate a plaintext version of the document. Specify FALSE to generate an HTML version of the document if you are using the AUTO_FILTER filter or indexing HTML documents.

**language**
Specify the language of the document. Use an Oracle Text supported language value as you would in the language column of the base table. See BASIC_LEXER on page 2-39 in Chapter 2, "Oracle Text Indexing Elements".

**format**
Specify the format of the document. Use an Oracle Text supported format value, either TEXT, BINARY or IGNORE as you would specify in the format column of the base table. For more information, see the format column description in CREATE_INDEX on page 1-36 in Chapter 1, "Oracle Text SQL Statements and Operators".

**charset**
Specify the character set of the document. Use an Oracle Text supported value as you would specify in the charset column of the base table. See Indexing Mixed-Character Set Columns on page 2-18 in Chapter 2, "Oracle Text Indexing Elements".
POLICY_GIST

Generates a Gist or theme summary for document. You can generate paragraph-level or sentence-level gists or theme summaries. With this procedure, no CONTEXT index is required.

**Note:** CTX_DOC.POLICY_GIST requires an installed knowledge base. A knowledge base may or may not have been installed with Oracle Text. For more information on knowledge bases, see *Oracle Text Application Developer’s Guide*.

### Syntax

```sql
ctx_doc.policy_gist(policy_name in VARCHAR2,
                    document in [VARCHAR2|CLOB|BLOB|BFILE],
                    restab in out nocopy CLOB,
                    glevel in VARCHAR2 default 'P',
                    pov in VARCHAR2 default 'GENERIC',
                    numParagraphs in VARCHAR2 default NULL,
                    maxPercent in NUMBER default NULL,
                    num_themes in NUMBER default 50,
                    language in VARCHAR2 default NULL,
                    format in VARCHAR2 default NULL,
                    charset in VARCHAR2 default NULL
);
```

**policy_name**
Specify the policy name created with CTX_DDL.CREATE_POLICY.

**document**
Specify the document for which to generate the Gist or theme summary.

**restab**
Specify the name of the CLOB locator.

**glevel**
Specify the type of gist or theme summary to produce. The possible values are:

- **P** for paragraph
- **S** for sentence

The default is **P**.

**pov**
Specify whether a gist or a single theme summary is generated. The type of gist or theme summary generated (sentence-level or paragraph-level) depends on the value specified for glevel.

To generate a gist for the entire document, specify a value of 'GENERIC' for pov. To generate a theme summary for a single theme in a document, specify the theme as the value for pov.

When using result table storage and you do not specify a value for pov, this procedure returns the generic gist plus up to 50 theme summaries for the document.
numParagraphs
Specify the maximum number of document paragraphs (or sentences) selected for the
document gist or theme summaries. The default is 16.

maxPercent
Specify the maximum number of document paragraphs (or sentences) selected for the
document gist or theme summaries as a percentage of the total paragraphs (or
sentences) in the document. The default is 10.

num_themes
Specify the number of theme summaries to produce when you do not specify a value
for pov. For example, if you specify 10, this procedure returns the top 10 theme
summaries. The default is 50.

If you specify 0 or NULL, this procedure returns all themes in a document. If the
document contains more than 50 themes, only the top 50 themes show conceptual
hierarchy.

language
Specify the language of the document. Use an Oracle Text supported language value
as you would in the language column of the base table. See "MULTI_LEXER" on
page 2-48.

format
Specify the format of the document. Use an Oracle Text supported format value, either
TEXT, BINARY or IGNORE as you would specify in the format column of the base
table. For more information, see the format column description in "CREATE INDEX"
on page 1-36.
charset
Specify the character set of the document. Use an Oracle Text supported value as you would specify in the charset column of the base table.
Generates plain text or HTML highlighting offset information for a document. With this procedure, no CONTEXT index is required.

The offset information is generated for the terms in the document that satisfy the query you specify. These highlighted terms are either the words that satisfy a word query or the themes that satisfy an ABOUT query.

You can generate highlight offsets for either plaintext or HTML versions of the document. You can apply the offset information to the same documents filtered with CTX_DOC.FILTER.

**Syntax**

```sql
exec ctx_doc.policy_highlight(
    policy_name  in  VARCHAR2,
    document     in  [VARCHAR2|CLOB|BLOB|BFILE],
    text_query   in VARCHAR2,
    restab       in out nocopy highlight_tab,
    plaintext    in boolean FALSE
    language     in VARCHAR2 default NULL,
    format       in VARCHAR2 default NULL,
    charset      in VARCHAR2 default NULL
);
```

```sql
exec ctx_doc.policy_highlight_clob_query(
    policy_name  in VARCHAR2,
    document     in [VARCHAR2|CLOB|BLOB|BFILE],
    text_query   in CLOB,
    restab       in out nocopy highlight_tab,
    plaintext    in boolean FALSE
    language     in VARCHAR2 default NULL,
    format       in VARCHAR2 default NULL,
    charset      in VARCHAR2 default NULL
);
```

**policy_name**
Specify the policy name created with CTX_DDL.CREATE_POLICY.

**document**
Specify the document to generate highlighting offset information.

**text_query**
Specify the original query expression used to retrieve the document. If NULL, no highlights are generated.

If `text_query` includes wildcards, stemming, or fuzzy matching which result in stopwords being returned, this procedure does not highlight the stopwords.

If `text_query` contains the threshold operator, the operator is ignored. This procedure always returns highlight information for the entire result set.

**restab**
Specify the name of the `highlight_tab` PL/SQL index-by-table type.

**See Also:** see "HIGHLIGHT" on page 8-10 for more information about the structure of the `highlight_tab` table type.
plaintext
Specify TRUE to generate a plaintext offsets of the document.
Specify FALSE to generate HTML offsets of the document if you are using the AUTO_FILTER filter or indexing HTML documents.

language
Specify the language of the document. Use an Oracle Text supported language value as you would in the language column of the base table. See "MULTI_LEXER" in Chapter 2, "Oracle Text Indexing Elements".

format
Specify the format of the document. Use an Oracle Text supported format value, either TEXT, BINARY or IGNORE as you would specify in the format column of the base table. For more information, see the format column description under "CREATE INDEX" on page 1-36.

charset
Specify the character set of the document. Use an Oracle Text supported value as you would specify in the charset column of the base table.

Restrictions
CTX_DOC.POLICY_HIGHLIGHT does not support the use of query templates.
Generates plain text or HTML version of a document with query terms highlighted. With this procedure, no CONTEXT index is required.

The CTX_Doc.POLICY_MARKUP procedure takes a query specification and a document and returns a version of the document in which the query terms are marked up. These marked-up terms are either the words that satisfy a word query or the themes that satisfy an ABOUT query.

You can set the marked-up output to be either plaintext or HTML.

You can use one of the pre-defined tag sets for marking highlighted terms, including a tag sequence that enables HTML navigation.

Syntax

```sql
ctx_doc.policy_markup(policy_name in VARCHAR2,
document        in [VARCHAR2|CLOB|BLOB|BFILE],
text_query      in VARCHAR2,
restab          in out nocopy CLOB,
plaintext       in BOOLEAN default FALSE,
tagset          in VARCHAR2 default 'TEXT_DEFAULT',
starttag        in VARCHAR2 default NULL,
endtag          in VARCHAR2 default NULL,
prevtag         in VARCHAR2 default NULL,
nexttag         in VARCHAR2 default NULL,
language        in VARCHAR2 default NULL,
charset         in VARCHAR2 default NULL
);

ctx_doc.policy_markup_clob_query(
policy_name     in VARCHAR2,
document        in [VARCHAR2|CLOB|BLOB|BFILE],
text_query      in CLOB,
restab          in out nocopy CLOB,
plaintext       in BOOLEAN default FALSE,
tagset          in VARCHAR2 default 'TEXT_DEFAULT',
starttag        in VARCHAR2 default NULL,
endtag          in VARCHAR2 default NULL,
prevtag         in VARCHAR2 default NULL,
nexttag         in VARCHAR2 default NULL,
language        in VARCHAR2 default NULL,
charset         in VARCHAR2 default NULL
);
```

`policy_name`
Specify the policy name created with CTX_DDL.CREATE_POLICY.

`document`
Specify the document to generate highlighting offset information.

`text_query`
Specify the original query expression used to retrieve the document.

If `text_query` includes a NULL, then this procedure will fail and generate errors.
If `text_query` includes wildcards, stemming, or fuzzy matching which result in stopwords being returned, then this procedure does not highlight the stopwords.

If `text_query` contains the threshold operator, the operator is ignored. This procedure always returns highlight information for the entire result set.

**restab**
Specify the name of the CLOB locator.

**plaintext**
Specify `TRUE` to generate a plaintext marked-up document. Specify `FALSE` to generate a marked-up HTML version of the document if you are using the `AUTO_FILTER` filter or indexing HTML documents.

**tagset**
Specify one of the following pre-defined tag sets. The second and third columns show how the four different tags are defined for each tagset:

<table>
<thead>
<tr>
<th>Tagset</th>
<th>Tag</th>
<th>Tag Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEXT_DEFAULT</td>
<td>starttag</td>
<td>&lt;&lt;&lt;</td>
</tr>
<tr>
<td></td>
<td>endtag</td>
<td>&gt;&gt;&gt;</td>
</tr>
<tr>
<td></td>
<td>prevtag</td>
<td></td>
</tr>
<tr>
<td></td>
<td>nexttag</td>
<td></td>
</tr>
<tr>
<td>HTML_DEFAULT</td>
<td>starttag</td>
<td>&lt;B&gt;</td>
</tr>
<tr>
<td></td>
<td>endtag</td>
<td>&lt;/B&gt;</td>
</tr>
<tr>
<td></td>
<td>prevtag</td>
<td></td>
</tr>
<tr>
<td></td>
<td>nexttag</td>
<td></td>
</tr>
<tr>
<td>HTML_NAVIGATE</td>
<td>starttag</td>
<td>&lt;A NAME=ctx%CURNUM&gt;&lt;B&gt;</td>
</tr>
<tr>
<td></td>
<td>endtag</td>
<td>&lt;/B&gt;&lt;/A&gt;</td>
</tr>
<tr>
<td></td>
<td>prevtag</td>
<td>&lt;A HREF=#ctx%PREVNUM&gt;&lt;&lt;/A&gt;</td>
</tr>
</tbody>
</table>
|                | nexttag | <A HREF=#ctx%NEXTNUM>&gt;</A>

**starttag**
Specify the character(s) inserted by `MARKUP` to indicate the start of a highlighted term.

The sequence of `starttag`, `endtag`, `prevtag` and `nexttag` with regard to the highlighted word is as follows:

... `prevtag starttag word endtag nexttag`...

**endtag**
Specify the character(s) inserted by `MARKUP` to indicate the end of a highlighted term.

**prevtag**
Specify the markup sequence that defines the tag that navigates the user to the previous highlight.

In the markup sequences `prevtag` and `nexttag`, you can specify the following offset variables which are set dynamically:
See the description of the HTML_NAVIGATE tagset for an example.

nexttag
Specify the markup sequence that defines the tag that navigates the user to the next highlight tag.

Within the markup sequence, you can use the same offset variables you use for prevtag. See the explanation for prevtag and the HTML_NAVIGATE tagset for an example.

language
Specify the language of the document. Use an Oracle Text supported language value as you would in the language column of the base table. See "MULTI_lexer" in Chapter 2, "Oracle Text Indexing Elements".

format
Specify the format of the document. Use an Oracle Text supported format value, either TEXT, BINARY or IGNORE as you would specify in the format column of the base table. For more information, see the format column description in "CREATE INDEX".

charset
Specify the character set of the document. Use an Oracle Text supported value as you would specify in the charset column of the base table. See "Indexing Mixed-Character Set Columns" in Chapter 2, "Oracle Text Indexing Elements".

Restrictions

CTX_DOC.POLICY_MARKUP does not support the use of query templates.
Display marked-up keywords in context. The returned text contains either the words that satisfy a word query or the themes that satisfy an ABOUT query. This version of the CTX_DOC.SNIPPET procedure does not require an index.

Syntax

Syntax 1
exec CTX_DOC.POLICY_SNIPPET(
policy_name              IN VARCHAR2,
document                 IN [VARCHAR2|CLOB|BLOB|BFILE],
text_query               IN VARCHAR2,
language                 IN VARCHAR2 default NULL,
charset                  IN VARCHAR2 default NULL,
starttag                 IN VARCHAR2 DEFAULT '<b>',
endtag                   IN VARCHAR2 DEFAULT '</b>',
extart_translation       IN BOOLEAN DEFAULT TRUE,
separator                IN VARCHAR2 DEFAULT '<b>...</b>'
) return varchar2;

Syntax 2
exec CTX_DOC.POLICY_SNIPPET_CLOB_QUERY(
policy_name              IN VARCHAR2,
document                 IN [VARCHAR2|CLOB|BLOB|BFILE],
text_query               IN VARCHAR2,
language                 IN VARCHAR2 default NULL,
charset                  IN VARCHAR2 default NULL,
starttag                 IN VARCHAR2 DEFAULT '<b>',
endtag                   IN VARCHAR2 DEFAULT '</b>',
extart_translation       IN BOOLEAN DEFAULT TRUE,
separator                IN VARCHAR2 DEFAULT '<b>...</b>'
) return varchar2;

**policy_name**
Specify the name of a policy created with CTX_DDL.CREATE_POLICY.

**document**
Specify the document in which to search for keywords.

**text_query**
Specify the original query expression used to retrieve the document. If NULL, no highlights are generated.

If text_query includes wildcards, stemming, fuzzy matching which result in stopwords being returned, POLICY_SNIPPET does not highlight the stopwords.

If text_query contains the threshold operator, the operator is ignored.
language
Specify the language of the document. Use an Oracle Text supported language value as you would in the language column of the base table. See MULTI_LEXER in Chapter 2, "Oracle Text Indexing Elements".

format
Specify the format of the document. Use an Oracle Text supported format value, either TEXT, BINARY or IGNORE as you would specify in the format column of the base table. For more information, see the format column description in "CREATE INDEX".

charset
Specify the character set of the document. Use an Oracle Text supported value as you would specify in the charset column of the base table. See "Indexing Mixed-Character Set Columns" in Chapter 2, "Oracle Text Indexing Elements".

starttag
Specify the start tag for marking up the query keywords. Default is '<b>'.

endtag
Specify the end tag for marking up the query keywords. Default is '</b>'.

entity_translation
Specify if you want HTML entities to be translated. The default is TRUE, which means the special entities ('<', '>', and '&') are translated into their alternate forms ('&lt;', '&gt;', and '&amp;') when output by the procedure. However, special characters in the markup tags generated by CTX_DOC.POLICY_SNIPPET will not be translated.

separator
Specify the string separating different returned fragments. Default is '<b>...</b>'.

Limitations
CTX_DOC.POLICY_SNIPPET does not support the use of query templates.

CTX_DOC.POLICY_SNIPPET displays marked-up keywords in context when used with NULL_SECTION_GROUP. However, there are limitations when using this procedure with XML documents. When used with XML_SECTION_GROUP or AUTO_SECTION_GROUP, the XML structure is ignored and user-specified tags are stripped out, which results in parts of surrounding text to be included in the returned snippet.

Related Topics
See Also: "SNIPPET" on page 8-36, "HIGHLIGHT" on page 8-10, and "MARKUP" on page 8-14
POLICY_THEMES

Generates a list of themes for a document. With this procedure, no CONTEXT index is required.

**Note:** CTX_DOC.POLICY_THEMES requires an installed knowledge base. A knowledge base may or may not have been installed with Oracle Text. For more information on knowledge bases, see the *Oracle Text Application Developer’s Guide*.

**Syntax**

```sql
ctx_doc.policy_themes(policy_name in VARCHAR2,
document       in [VARCHAR2|CLOB|BLOB|BFILE],
restab         in out nocopy theme_tab,
full_themes    in BOOLEAN default FALSE,
num_themes     in number    default 50
language       in VARCHAR2 default NULL,
format         in VARCHAR2 default NULL,
charset        in VARCHAR2 default NULL);
```

**policy_name**  
Specify the policy you create with CTX_DDL.CREATE_POLICY.

**document**  
Specify the document for which to generate a list of themes.

**restab**  
Specify the name of the theme_tab PL/SQL index-by-table type.

**full_themes**  
Specify whether this procedure generates a single theme or a hierarchical list of parent themes (full themes) for each document theme.

**See Also:** "THEMES" on page 8-39 for more information about the structure of the theme_tab type.

**full_themes**  
Specify whether this procedure generates a single theme or a hierarchical list of parent themes (full themes) for each document theme.

Specify **TRUE** for this procedure to write full themes to the THEME column of the result table.

Specify **FALSE** for this procedure to write single theme information to the THEME column of the result table. This is the default.

**num_themes**  
Specify the maximum number of themes to retrieve. For example, if you specify 10, up to first 10 themes are returned for the document. The default is 50.

If you specify 0 or NULL, this procedure returns all themes in a document. If the document contains more than 50 themes, only the first 50 themes show conceptual hierarchy.

**language**  
Specify the language of the document. Use an Oracle Text supported language value as you would in the language column of the base table. See "MULTI_LEXER" in Chapter 2, "Oracle Text Indexing Elements".
**format**
Specify the format of the document. Use an Oracle Text supported format value, either TEXT, BINARY or IGNORE as you would specify in the format column of the base table. For more information, see the format column description in "CREATE INDEX" in Chapter 1, "Oracle Text SQL Statements and Operators".

**charset**
Specify the character set of the document. Use an Oracle Text supported value as you would specify in the charset column of the base table. See "Indexing Mixed-Character Set Columns" in Chapter 2, "Oracle Text Indexing Elements".

**Example**
Create a policy:
```
exec ctx_ddl.create_policy('mypolicy');
```
Run themes:
```
declare
  la     varchar2(200);
  rtab   ctx_doc.theme_tab;
begin
  ctx_doc.policy_themes('mypolicy',
    'To define true madness, What is''t but to be nothing but mad?', rtab);
  for i in 1..rtab.count loop
    dbms_output.put_line(rtab(i).theme||'':'||rtab(i).weight);
  end loop;
end;
```
Generate all index tokens for document. With this procedure, no CONTEXT index is required.

Syntax

```sql
ctx_doc.policy_tokens(policy_name in VARCHAR2,
document in [VARCHAR2|CLOB|BLOB|BFILE],
restab in out nocopy token_tab,
language in VARCHAR2 default NULL,
format in VARCHAR2 default NULL,
charset in VARCHAR2 default NULL);
```

**policy_name**
Specify the policy name created with CTX_DDL.CREATE_POLICY.

**document**
Specify the document for which to generate tokens.

**restab**
Specify the name of the `token_tab` PL/SQL index-by-table type.

The tokens returned are those tokens which are inserted into the index for the document. Stop words are not returned. Section tags are not returned because they are not text tokens.

**See Also:** "TOKENS" on page 8-42 of this chapter for more information about the structure of the `token_tab` type

**language**
Specify the language of the document. Use an Oracle Text supported language value as you would in the language column of the base table. See "MULTI_LEXER" in Chapter 2, "Oracle Text Indexing Elements".

**format**
Specify the format of the document. Use an Oracle Text supported format value, either TEXT, BINARY or IGNORE as you would specify in the format column of the base table. For more information, see the format column description in "CREATE INDEX".

**charset**
Specify the character set of the document. Use an Oracle Text supported value as you would specify in the charset column of the base table. See "Indexing Mixed-Character Set Columns" in Chapter 2, "Oracle Text Indexing Elements".

**Example**

Get tokens:

```sql
declare
  la     varchar2(200);
  rtab   ctx_doc.token_tab;
begin
  ctx_doc.policy_tokens('mypolicy',
    'To define true madness, What is''t but to be nothing but mad?',rtab);
  for i in 1..rtab.count loop
    dbms_output.put_line(rtab(i).offset||':'||rtab(i).token);
  end loop;
end;
```

end loop;
end;
Use this procedure to set the CTX_DOC procedures to accept either the ROWID or the PRIMARY_KEY document identifiers. This setting affects the invoking session only.

**Syntax**

```sql
ctx_doc.set_key_type(key_type in varchar2);
```

**key_type**

Specify either ROWID or PRIMARY_KEY as the input key type (document identifier) for CTX_DOC procedures.

This parameter defaults to the value of the CTX_DOC_KEY_TYPE system parameter.

**Note:** When your base table has no primary key, setting key_type to PRIMARY_KEY is ignored. The textkey parameter that you specify for any CTX_DOC procedure is interpreted as a ROWID.

**Example**

The following example sets CTX_DOC procedures to accept primary key document identifiers.

```sql
begin
ctx_doc.set_key_type('PRIMARY_KEY');
end
```
Use the `CTX_DOC.SNIPPET` procedure to produce a concordance for a document. A concordance is a text fragment that contains a query term with some of its surrounding text. This is also sometimes known as Key Word in Context or KWIC, because it returns query keywords marked up in their surrounding text, which enables the user to evaluate them in context. The returned text can also contain themes that satisfy an `ABOUT` query.

For example, a search on `brillig` and `slithey` might return one relevant fragment of a document as follows:

'Twas <b>brillig</b>, and the <b>slithey</b> toves did gyre and

`CTX_DOC.SNIPPET` returns one or more most relevant fragments for a document that contains the query term. Because `CTX_DOC.SNIPPET` returns surrounding text, you can immediately evaluate how useful the returned term is. `CTX_DOC.SNIPPET` returns the entire document if no words in the returned text are marked up.

See Also: `CTX_DOC.POLICY_SNIPPET` on page 8-29 for a policy-based version of this procedure

Syntax

**Syntax 1**
```sql
exec CTX_DOC.SNIPPET(
    index_name              IN VARCHAR2,
    textkey                  IN VARCHAR2,
    text_query               IN VARCHAR2,
    starttag                 IN VARCHAR2 DEFAULT '<b>',
    endtag                   IN VARCHAR2 DEFAULT '</b>',
    entity_translation       IN BOOLEAN  DEFAULT TRUE,
    separator                IN VARCHAR2 DEFAULT '<b>...</b>'
) return varchar2;
```

**Syntax 2**
```sql
exec CTX_DOC.SNIPPET_CLOB_QUERY(
    index_name               IN VARCHAR2,
    textkey                  IN CLOB,
    text_query               IN VARCHAR2,
    starttag                 IN VARCHAR2 DEFAULT '<b>',
    endtag                   IN VARCHAR2 DEFAULT '</b>',
    entity_translation       IN BOOLEAN  DEFAULT TRUE,
    separator                IN VARCHAR2 DEFAULT '<b>...</b>'
) return varchar2;
```

**index_name**
Specify the name of the index for the text column.

**textkey**
Specify the unique identifier (usually the primary key) for the document.

The `textkey` parameter can be as follows:

- A single column primary key value
- An encoded specification for a composite (multiple column) primary key. When textkey is a composite key, you must encode the composite textkey string using the CTX_DOC.PKENCODE procedure.

- The rowid of the row containing the document

Use CTX_DOC.SET_KEY_TYPE to toggle between primary key and rowid identification.

text_query
Specify the original query expression used to retrieve the document. If NULL, no highlights are generated.

If text_query includes wildcards, stemming, fuzzy matching which result in stopwords being returned, SNIPPET does not highlight the stopwords.

If text_query contains the threshold operator, the operator is ignored.

starttag
Specify the start tag for marking up the query keywords. Default is '<b>'.

dentag
Specify the end tag for marking up the query keywords. Default is '</b>'.

entity_translation
Specify if you want HTML entities to be translated. The default is TRUE, which means that the special entities ('<', '>', and '&') are translated into their alternative forms ('&lt;', '&gt;', and '&amp;') when output by the procedure. However, special characters in the markup tags that are generated by CTX_DOC.SNIPPET will not be translated.

separator
Specify the string separating different returned fragments. Default is '<b>...</b>'.
The result looks something like this:

```
CTX_DOC.SNIPPET('TDRBHKO1X', '1', 'SEARCH|CLASSIFY')
```

Text's unique knowledge-base enables you to <b>search</b>, <b>classify</b>, manage documents, clusters and summarize.

**Limitations**

CTX_DOC.SNIPPET does not support the use of query templates.

CTX_DOC.SNIPPET displays marked-up keywords in context when used with NULL_SECTION_GROUP. However, there are limitations when using this procedure with XML documents. When used with XML_SECTION_GROUP or AUTO_SECTION_GROUP, the XML structure is ignored and user-specified tags are stripped out, which results in parts of surrounding text to be included in the returned snippet.

**Related Topics**

See Also: "POLICY_SNIPPET" on page 8-29, "HIGHLIGHT" on page 8-10, and "MARKUP" on page 8-14
THEMES

Use the CTX_DOC.THEMES procedure to generate a list of themes for a document. You can store each theme as a row in either a result table or an in-memory PL/SQL table that you specify.

Note: CTX_DOC.THEMES requires an installed knowledge base. A knowledge base may or may not have been installed with Oracle Text. For more information on knowledge bases, see Oracle Text Application Developer's Guide.

Syntax 1: In-Memory Table Storage

    CTX_DOC.THEMES(
        index_name      IN VARCHAR2,
        textkey         IN VARCHAR2,
        restab          IN OUT NOCOPY THEME_TAB,
        full_themes     IN BOOLEAN DEFAULT FALSE,
        num_themes       IN NUMBER DEFAULT 50);

Syntax 2: Result Table Storage

    CTX_DOC.THEMES(
        index_name      IN VARCHAR2,
        textkey         IN VARCHAR2,
        restab          IN VARCHAR2,
        query_id        IN NUMBER DEFAULT 0,
        full_themes     IN BOOLEAN DEFAULT FALSE,
        num_themes       IN NUMBER DEFAULT 50);

  index_name
Specify the name of the index for the text column.

  textkey
Specify the unique identifier (usually the primary key) for the document.

  The textkey parameter can be as follows:
  
  ■ A single column primary key value
  
  ■ An encoded specification for a composite (multiple column) primary key. When textkey is a composite key, you must encode the composite textkey string using the CTX_DOC.PKENCODE procedure.

  ■ The rowid of the row containing the document

  Toggle between primary key and rowid identification using CTX_DOC.SET_KEY_TYPE.

  restab
You can specify this procedure to store results to either a table or to an in-memory PL/SQL table.

  To store results in a table, specify the name of the table.

  See Also: "Theme Table" on page A-7 in Appendix A, "Oracle Text Result Tables"
To store results in an in-memory table, specify the name of the in-memory table of type `THEME_TAB`. The `THEME_TAB` datatype is defined as follows:

```sql
type theme_rec is record (
    theme varchar2(2000),
    weight number
);

type theme_tab is table of theme_rec index by binary_integer;
```

`CTX_DOC.THEMES` clears the `THEME_TAB` you specify before the operation.

**query_id**
Specify the identifier used to identify the row(s) inserted into `restab`.

**full_themes**
Specify whether this procedure generates a single theme or a hierarchical list of parent themes (full themes) for each document theme.

Specify `TRUE` for this procedure to write full themes to the `THEME` column of the result table.

Specify `FALSE` for this procedure to write single theme information to the `THEME` column of the result table. This is the default.

**num_themes**
Specify the maximum number of themes to retrieve. For example, if you specify 10, then up to the first 10 themes are returned for the document. The default is 50.

If you specify 0 or `NULL`, then this procedure returns all themes in a document. If the document contains more than 50 themes, then only the first 50 themes show conceptual hierarchy.

### Examples

**In-Memory Themes**
The following example generates the first 10 themes for document 1 and stores them in an in-memory table called `the_themes`. The example then loops through the table to display the document themes.

```sql
declare
    the_themes ctx_doc.theme_tab;
begin
    ctx_doc.themes('myindex','1',the_themes, num_themes=>10);
    for i in 1..the_themes.count loop
        dbms_output.put_line(the_themes(i).theme||':'||the_themes(i).weight);
    end loop;
end;
```

**Theme Table**
The following example creates a theme table called `CTX_THEMES`:

```sql
create table CTX_THEMES (query_id number,
    theme varchar2(2000),
    weight number);
```
**Single Themes**

To obtain a list of up to the first 20 themes, where each element in the list is a single theme, enter a statement like the following example:

```plaintext
begin
  ctx_doc.themes('newsindex','34','CTX_THEMES',1,full_themes => FALSE,
              num_themes=> 20);
end;
```

**Full Themes**

To obtain a list of the top 20 themes, where each element in the list is a hierarchical list of parent themes, enter a statement like the following example:

```plaintext
begin
  ctx_doc.themes('newsindex','34','CTX_THEMES',1,full_themes => TRUE, num_themes=>20);
end;
```
TOKENS

Use this procedure to identify all text tokens in a document. The tokens returned are those tokens that are inserted into the index. This feature is useful for implementing document classification, routing, or clustering.

Stopwords are not returned. Section tags are not returned because they are not text tokens.

Syntax 1: In-Memory Table Storage

```sql
CTX_DOC.TOKENS(index_name IN VARCHAR2,
    textkey IN VARCHAR2,
    restab IN OUT NOCOPY TOKEN_TAB);
```

Syntax 2: Result Table Storage

```sql
CTX_DOC.TOKENS(index_name IN VARCHAR2,
    textkey IN VARCHAR2,
    restab IN VARCHAR2,
    query_id IN NUMBER DEFAULT 0);
```

**index_name**
Specify the name of the index for the text column.

**textkey**
Specify the unique identifier (usually the primary key) for the document.

The `textkey` parameter can be as follows:

- A single column primary key value
- Encoded specification for a composite (multiple column) primary key. To encode a composite `textkey`, use the `CTX_DOC.PKENCODE` procedure.
- The rowid of the row containing the document

Toggle between primary key and rowid identification using `CTX_DOC.SET_KEY_TYPE`.

**restab**
You can specify that this procedure store results to either a table or to an in-memory PL/SQL table.

The tokens returned are those tokens that are inserted into the index for the document (or row) named with `textkey`. Stop words are not returned. Section tags are not returned because they are not text tokens.

**Specifying a Token Table**

To store results to a table, specify the name of the table. Token tables can be named anything, but must include the columns shown in the following table, with names and datatypes as specified.
Specifying an In-Memory Table
To store results to an in-memory table, specify the name of the in-memory table of type `TOKEN_TAB`. The `TOKEN_TAB` datatype is defined as follows:

```sql
type token_rec is record (
  token varchar2(64),
  offset number,
  length number
);

type token_tab is table of token_rec index by binary_integer;

CTX_DOC.TOKENS clears the `TOKEN_TAB` you specify before the operation.
```

**query_id**
Specify the identifier used to identify the row(s) inserted into restab.

### Example

**In-Memory Tokens**
The following example generates the tokens for document 1 and stores them in an in-memory table, declared as `the_tokens`. The example then loops through the table to display the document tokens.

```sql
declare
  the_tokens ctx_doc.token_tab;

begin
  ctx_doc.tokens('myindex','1',the_tokens);
  for i in 1..the_tokens.count loop
    dbms_output.put_line(the_tokens(i).token);
  end loop;
end;
```
This chapter provides reference information for using the CTX_ENTITY PL/SQL package. This package is used to locate and classify words and phrases into categories, such as persons or companies.

CTX_ENTITY contains the following stored procedures and functions.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD_EXTRACT_RULE</td>
<td>Adds a single extraction rule to an extraction policy.</td>
</tr>
<tr>
<td>ADD_STOP_ENTITY</td>
<td>Marks certain entity mentions or entity types as not to be extracted.</td>
</tr>
<tr>
<td>COMPILIE</td>
<td>Compiles added extraction rules into an extraction policy.</td>
</tr>
<tr>
<td>CREATE_EXTRACT_POLICY</td>
<td>Creates an extraction policy to use.</td>
</tr>
<tr>
<td>DROP_EXTRACT_POLICY</td>
<td>Drops an extraction policy.</td>
</tr>
<tr>
<td>EXTRACT</td>
<td>Generates an XML document describing the entities found in an input document.</td>
</tr>
<tr>
<td>REMOVE_EXTRACT_RULE</td>
<td>Removes a single extraction rule from an extraction policy.</td>
</tr>
<tr>
<td>REMOVE_STOP_ENTITY</td>
<td>Removes a stop entity from an extraction policy.</td>
</tr>
</tbody>
</table>
This procedure adds a single extraction rule to extract policy. Invokers add rules into their own extraction policy. Extraction rules have sentence-wide scopes. Extraction rules have to be case-sensitive except for entity types and rule operators in the rule expression. Order of rule addition is not important. Addition of a rule will not be effective until CTX_ENTITY.COMPILE is executed. This procedure issues a commit.

Syntax

```
CTX_ENTITY.ADD_EXTRACT_RULE(
    policy_name                 IN VARCHAR2,
    rule_id                     IN INTEGER,
    extraction_rule             IN VARCHAR2);
```

**policy_name**
Specify the policy name.

**rule_id**
Specify a unique rule ID within an extraction policy. The rule ID must be greater than 0.

**extraction_rule**
The rule text in XML format specifies the language, expression, and entities to be extracted. The rule text follows the XML schema below:

```xml
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="rule">
    <xsd:sequence>
      <xsd:element name="expression" type="xsd:string"/>
      <xsd:complexType>
        <xsd:attribute name="refid" type="xsd:positiveInteger"/>
      </xsd:complexType>
      <xsd:element name="comments" type="xsd:string" default="\0"/>
    </xsd:sequence>
  </xsd:element>
</xsd:schema>
```

Where:

- The language attribute of the rule tag specifies the applied language for the rule. The rule will only be applied to documents that are of the specified languages. The language attribute can be left out, or set to "ALL" if the rule is to match on all documents.
- The expression tag contains the posix regular expression that will be used in the matching.
- The comments tag allows users to associate any comments with this user rule.
- The type tag assigns the extracted entity text to a given entity type. The refid attribute of the type tag specifies which backreference in the regular expression corresponds to the actual entity. The entity type can be one of the Oracle supplied types, listed in Table 9–1, "Supplied Entity Types", or it can be a user-defined type, which must be prefixed with the letter "x". 

The following example shows how to define an extraction rule and associate it with an entity extraction policy. The following rule defines a simple extraction rule for finding email addresses in documents.

begin
ctx_entity.add_extract_rule('pol1', 1,
'<rule>
  <expression>email is \w+@\w+\.\w+</expression>
</rule>
')

Example 1

Table 9–1  Supplied Entity Types

<table>
<thead>
<tr>
<th>Supplied Entity Type</th>
<th>Explanation</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>building</td>
<td>A particular building</td>
<td>White House</td>
</tr>
<tr>
<td>city</td>
<td></td>
<td>New York</td>
</tr>
<tr>
<td>company</td>
<td></td>
<td>Oracle Corporation</td>
</tr>
<tr>
<td>country</td>
<td></td>
<td>United States</td>
</tr>
<tr>
<td>currency</td>
<td></td>
<td>Dollar</td>
</tr>
<tr>
<td>date</td>
<td></td>
<td>July 4</td>
</tr>
<tr>
<td>day</td>
<td></td>
<td>Monday, Tuesday</td>
</tr>
<tr>
<td>email_address</td>
<td></td>
<td><a href="mailto:scott.tiger@oracle.com">scott.tiger@oracle.com</a></td>
</tr>
<tr>
<td>geo_political</td>
<td>A political or strategic</td>
<td>United Nations</td>
</tr>
<tr>
<td>holiday</td>
<td></td>
<td>Labor Day</td>
</tr>
<tr>
<td>location_other</td>
<td>Other types of locations</td>
<td>Atlantic Ocean</td>
</tr>
<tr>
<td>month</td>
<td></td>
<td>June, July</td>
</tr>
<tr>
<td>non_profit</td>
<td>Non-profit organization</td>
<td>Red Cross</td>
</tr>
<tr>
<td>organization_other</td>
<td>Other types of organizations</td>
<td>Supreme Court</td>
</tr>
<tr>
<td>percent</td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>person_jobtitle</td>
<td>Person referred to by title</td>
<td>President, Professor</td>
</tr>
<tr>
<td>person_name</td>
<td>Person referred to by name</td>
<td>John Doe</td>
</tr>
<tr>
<td>person_other</td>
<td>Other types of persons</td>
<td>Other types of persons (for example, criminal)</td>
</tr>
<tr>
<td>phone_number</td>
<td></td>
<td>(123)-456-7890</td>
</tr>
<tr>
<td>postal_address</td>
<td></td>
<td>Redwood Shores, CA</td>
</tr>
<tr>
<td>product</td>
<td></td>
<td>Oracle Text</td>
</tr>
<tr>
<td>region</td>
<td></td>
<td>North America</td>
</tr>
<tr>
<td>ssn</td>
<td>Social Security Number</td>
<td>123-45-6789</td>
</tr>
<tr>
<td>state</td>
<td>A state or province</td>
<td>California</td>
</tr>
<tr>
<td>time_duration</td>
<td>A length of time</td>
<td>10 seconds</td>
</tr>
<tr>
<td>tod</td>
<td>Time of day</td>
<td>8:00 AM</td>
</tr>
<tr>
<td>url</td>
<td>Web address</td>
<td><a href="http://www.oracle.com">www.oracle.com</a></td>
</tr>
<tr>
<td>zip_code</td>
<td>Zip Code</td>
<td>CA 94065</td>
</tr>
</tbody>
</table>
Example 2

The following rule defines a simple extraction rule for finding phone numbers in documents:

```
begin
  ctx_entity.add_extract_rule('pol1', 2,
    '<rule language="english">
      <expression>(\(d{3}\) \d{3}-\d{3}-\d{4})</expression>
      <comments>Rule for phone numbers</comments>
      <type refid="1">phone_number</type>
    </rule>'
  end;
end;
```

Where:

- Given the sentence: "I can be contacted at (123) 456-7890", this extraction rule will extract "(123) 456-7890" as an entity of type phone_number.
- The rule is added to the extraction policy called pol1.
- The rule is added with rule ID of 2.
- The XML description of the rule is as follows:
  - The language attribute of the rule tag is set to english, so the rule will only apply to English documents.
  - The expression tag contains the regular expression to use in the extraction.
  - The value of the type element and the refid attribute of the type tag specify that the first backreference corresponds to the text of the entity.
  - Explanatory comments are associated with this rule.
**ADD_STOP_ENTITY**

This procedure is used to mark certain entity mentions or entity types as not to be extracted. Invokers add stop entities to their own extraction policy. It does not take effect until after CTX_ENTITY.COMPILE is run. Either entity_name or entity_type can be NULL, but not both. If one stop entity is a subset of another, it will be marked as a subset after CTX_ENTITY.COMPILE, and not used in extraction. This procedure issues a commit.

**Syntax**

```
CTX_ENTITY.ADD_STOP_ENTITY(
    policy_name                 IN VARCHAR2,
    entity_name                 IN INTEGER,
    entity_type                 IN VARCHAR2 DEFAULT NULL,
    comments                    IN VARCHAR2 DEFAULT NULL);
```

- **policy_name**
  Specify the policy name of the stop entity that is to be added.

- **entity_name**
  Specify the entity name to be listed as a stop entity. If entity_type is NULL, all mentions with this entity_name will be listed as stop entities. It is case-sensitive.

- **entity_type**
  If entity_name is NULL, this will specify an entire entity type to be listed as stop entity. If entity_name is not NULL, this will specify only the mention <entity_type, entity_name> as a stop entity. It is case-insensitive. The maximum byte length is 4000 bytes.

- **comments**
  The maximum byte length is 4000 bytes.

**Example**

The following adds a stop entity corresponding to all persons. After compilation, extraction will not report any mentions of entity type person.
```
exec ctx_entity.add_stop_entity('poll', NULL, 'person');
```

The following adds a stop entity corresponding to <'person', 'john doe'>. After compilation, extraction will not report any mentions of the pair <'person', 'john doe'>. This stop entity is actually a subset of the first stop entity added. It will be marked subset in the CTX_USER_EXTRACT_STOP_ENTITIES view, and will not be used in extraction.
```
exec ctx_entity.add_stop_entity('poll', 'john doe', 'person');
```

The following adds a stop entity corresponding to all mentions of ford. After compilation, extraction will not report any mentions of the entity ford, irrespective of the entity type of the mention. For example, if a rule matches ford to a person, the extraction will not report this match. If a rule matches ford to a company, the extraction will again not report this match.
```
exec ctx_entity.add_stop_entity('poll', 'ford', NULL);
```
This procedure compiles added extraction rules into an extraction policy. It can also be used to compile added stop entities into an extraction policy. Users have to invoke this procedure if they have added any rules or stop entities to their policy.

Invokers compile rules and stop entities into their own extraction policy. Users can choose to compile added rules, added stop entities, or both.

After compilation, the CTX_USER_EXTRACT_RULES and CTX_USER_EXTRACT_STOP_ENTITIES views will show which rules and stop entities are being used in the entity extraction.

Syntax

CTX_ENTITY.COMPILE(
   policy_name              IN VARCHAR2,
   compile_choice           IN NUMBER DEFAULT COMPILE_ALL,
   locking                  IN NUMBER DEFAULT LOCK_NOWAIT_ERROR);

policy_name
Specify the policy name that is to be compiled.

compile_choice
Specify the entity name to be listed as a stop entity. If entity_type is NULL, all mentions with this entity_name will be listed as stop entities. It is case-sensitive.

The options are COMPILE_ALL, COMPILE_RULES, and COMPILE_STOP_ENTITIES. COMPILE_ALL compiles both rules and stop entities. COMPILE_RULES compiles only rules. COMPILE_STOP_ENTITIES compiles only stop entities.

locking
The maximum byte length is 4000 bytes. Configure how COMPIL e deals with the situation where another COMPIL e is already running on the same policy.

The options for locking are:

- CTX_ENTITY.LOCK_WAIT
  If another compile is running, wait until the running compile is complete, then begin compile. (In the event of not being able to get a lock, it will wait forever and ignore the maxtime setting.).

- CTX_ENTITY.LOCK_NOWAIT
  If another compile is running, immediately returns without error.

- CTX_ENTITY.LOCK_NOWAIT_ERROR
  If another sync is running, error "DRG-51313: timeout while waiting for DML or optimize lock" is raised.

Example

The following compiles the policy using the default setting:

exec ctx_entity.compile('pol1');

The following compiles only the stop entities for the policy:

exec ctx_entity.compile('pol1', CTX_ENTITY.COMPILE_STOP_ENTITIES);
The following compiles both rules and stop entities. If a lock exists, the function returns immediately, but does not raise an error.

```python
def ctx_entity.compile('pol1', CTX_ENTITY.COMPILE_ALL,
                      CTX_ENTITY.LOCK_NOWAIT);
```
CREATE_EXTRACT_POLICY

This procedure creates an extraction policy to use. This policy can only be used by the policy owner.

Syntax

```sql
CTX_ENTITY.CREATE_EXTRACT_POLICY(
    policy_name                   IN VARCHAR2,
    lexer                         IN VARCHAR2 DEFAULT NULL,
    include_supplied_rules        IN BOOLEAN DEFAULT TRUE,
    include_supplied_dictionary   IN BOOLEAN DEFAULT TRUE);
```

- **policy_name**
  Specify the name of the new extraction policy.

- **lexer**
  Specify the name of the lexer preference. Only `auto_lexer` is supported. If not specified, `CTXSYS.DEFAULT_EXTRACT_LEXER` will be used. The attributes `index_stems` and `deriv_stems` are not allowed.

- **include_supplied_rules**
  Specify whether Oracle-supplied rules are included in entity extraction. If false, automatic acronym resolution will be turned off. The default is `true`.

- **include_supplied_dictionary**
  Specify whether the Oracle-supplied dictionary is included in entity extraction. The default is `true`.

Examples

The following creates an extraction policy using the default settings. By default, the Oracle-supplied features, such as rules and dictionary, are enabled.

```sql
exec ctx_entity.create_extract_policy('pol1');
```

The following creates an extraction policy that explicitly specifies certain parameters. It specifies the lexer to be used as `mylex`, which must be an autolexer preference. It also includes the Oracle-supplied rules but disables the Oracle-supplied dictionary.

```sql
exec ctx_entity.create_extract_policy('pol2', 'mylex', TRUE, FALSE);
```
This procedure drops an extraction policy. These policies can only be dropped by the policy owner. This procedure issues a commit.

Syntax

```plaintext
CTX_ENTITY.DROP_EXTRACT_POLICY(
    policy_name     IN VARCHAR2);
```

**policy_name**
Specify the name of the extraction policy to be dropped.

Example

The following drops the extraction policy `pol2`:

```plaintext
exec ctx_entity.drop_extract_policy('pol2');
```
This procedure runs entity extraction on a given document and generates an XML document describing the entities found in the document. The XML document will give the entity text, type, and location of the entity in the document. The extraction will use the settings (rules, stop entities, and dictionary) defined in the given extraction policy.

Entity type names in the result will be uppercased. Invokers can run extraction using their own extraction policy.

Before execution, you have to issue CTX_ENTITY.COMPILE.

**Syntax**

```plaintext
CTX_ENTITY.EXTRACT(
    policy_name            IN VARCHAR2,
    document               IN CLOB,
    language               IN VARCHAR2,
    result                  IN OUT NOCOPY CLOB,
    entity_type_list        IN CLOB DEFAULT NULL);
```

**policy_name**

Run extraction using the given policy.

**document**

The input document to run extraction on.

If `entity_type` is NULL, all mentions with this `entity_name` will be listed as stop entities. It is case-sensitive.

**language**

Specify the language of the input document. If no language is specified, the language will be auto-detected.

The maximum byte length is 4000 bytes.

**result**

A CLOB containing the XML description of the entities extracted from the document.

If `entity_type` is NULL, all mentions with this `entity_name` will be listed as stop entities. It is case-sensitive.

**entity_type_list**

Specify that extraction will only consider a subset of entity types. The `entity_type_list` is a comma-separated list. If the `entity_type_list` is not specified, the entity extraction will consider all entity types.

**Example**

The following example shows the results of entity extraction on an example document. Suppose that we have created an extraction policy called `pol1`, and we are given the input document:

```
Sam A. Schwartz retired as executive vice president of Hupplewhite INc. in New York.
```

We then call the `ctx_entity.extract` procedure to generate an XML document containing the entities in this document. We insert the results CLOB into a table called `entities` for future viewing.
declare
    myresults clob;
begin
    select txt into mydoc from docs where id=1;
    ctx_entity.extract('pl', mydoc, null, myresults);
    insert into entities values(1, myresults);
    commit;
end;
/

Then we can examine the extracted entities from the entities table. Note that each entity is tagged with its location in the input document, as well as the source used to classify the entity.

<entities>
<entity id="0" offset="75" length="8" source="SuppliedDictionary">
    <text>New York</text>
    <type>city</type>
</entity>
<entity id="1" offset="55" length="16" source="SuppliedRule">
    <text>Hupplewhite Inc.</text>
    <type>company</type>
</entity>
<entity id="2" offset="27" length="24" source="SuppliedDictionary">
    <text>Sam A. Schwartz</text>
    <type>person_name</type>
</entity>
<entity id="4" offset="75" length="8" source="SuppliedDictionary">
    <text>New York</text>
    <type>state</type>
</entity>
</entities>
This procedure removes an extraction rule from the specified policy given a rule_id. Only the owner of the specified policy can remove an extraction rule from the policy. Removal of the extraction rule will be in effect after running CTX_ENTITY.COMPILE.

Syntax

```sql
CTX_ENTITY.REMOVE_EXTRACT_RULE(
    policy_name       IN VARCHAR2,
    rule_id           IN INTEGER);
```

case_name
Remove the extraction rule from the specified policy.

rule_id
Specify the rule ID of the extraction rule to be removed.

Example

The following removes the extraction rule with ID 1 from the policy pol1:

```sql
exec ctx_entity.remove_extract_rule('pol1', 1);
```
REMOVE_STOP_ENTITY

This procedure removes a stop entity from an extraction policy. Only the owner of the specified policy can remove a stop entity from the policy. Removal of the stop entity will be in effect after running CTX_ENTITY.COMPILE. Either the entity_name or entity_type can be null, but not both.

Syntax

```java
CTX_ENTITY.REMOVE_STOP_ENTITY(
    policy_name                 IN VARCHAR2,
    entity_name                 IN INTEGER DEFAULT NULL,
    entity_type                 IN VARCHAR2 DEFAULT NULL);
```

- **policy_name**
  Remove the stop_entity from the specified policy.

- **entity_name**
  Specify the name to be removed from the stop entity list. The stop_entity must have already been added to the stop_entity list using CTX_ENTITY.ADD_STOP_ENTITY.

- **entity_type**
  Specify the type of entity to be removed from the stop entity list. The stop_entity must have already been added to the stop entity list using CTX_ENTITY.ADD_STOP_ENTITY.

Example

```sql
exec ctx_entity.remove_stop_entity('pol1', NULL, 'person_name');
```

The above statement removes the stop entity corresponding to all mentions of the entity_type person_name from the policy pol1. After execution, this stop entity will be marked as "to be deleted" in the CTX_USER_EXTRACT_STOP_ENTITIES view. The removal of the stop entity will take effect once the user runs CTX_ENTITY.COMPILE.
This chapter provides reference information for using the \texttt{CTX\_OUTPUT} PL/SQL package.

\texttt{CTX\_OUTPUT} contains the following stored procedures:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{ADD_EVENT}</td>
<td>Adds an event to the index log.</td>
</tr>
<tr>
<td>\texttt{ADD_TRACE}</td>
<td>Enables tracing.</td>
</tr>
<tr>
<td>\texttt{DISABLE_QUERY_STATS}</td>
<td>Turns off the gathering of query stats for the index.</td>
</tr>
<tr>
<td>\texttt{ENABLE_QUERY_STATS}</td>
<td>Enables gathering of query stats for the index.</td>
</tr>
<tr>
<td>\texttt{END_LOG}</td>
<td>Halts logging of index and document services requests.</td>
</tr>
<tr>
<td>\texttt{END_QUERY_LOG}</td>
<td>Stops logging queries into a logfile.</td>
</tr>
<tr>
<td>\texttt{GET_TRACE_VALUE}</td>
<td>Returns the value of a trace.</td>
</tr>
<tr>
<td>\texttt{LOG_TRACES}</td>
<td>Prints traces to logfile.</td>
</tr>
<tr>
<td>\texttt{LOGFILENAME}</td>
<td>Returns the name of the current log file.</td>
</tr>
<tr>
<td>\texttt{REMOVE_EVENT}</td>
<td>Removes an event from the index log.</td>
</tr>
<tr>
<td>\texttt{REMOVE_TRACE}</td>
<td>Disables tracing.</td>
</tr>
<tr>
<td>\texttt{RESET_TRACE}</td>
<td>Clears a trace.</td>
</tr>
<tr>
<td>\texttt{START_LOG}</td>
<td>Starts logging index and document service requests.</td>
</tr>
<tr>
<td>\texttt{START_QUERY_LOG}</td>
<td>Creates a log file of queries.</td>
</tr>
</tbody>
</table>
ADD EVENT

Use this procedure to add an event to the index log for more detailed log output or to enable error tracing for Oracle Text errors.

Syntax

```sql
CTX_OUTPUT.ADD_EVENT(event in NUMBER, errnum in NUMBER := null);
```

**event**
Specify the type of index event to log. You can add the following events:

- `CTX_OUTPUT.EVENT_INDEX_PRINT_ROWID`, which logs the rowid of each row after it is indexed. This is useful for debugging a failed index operation.
- `CTX_OUTPUT.EVENT_OPT_PRINT_TOKEN`, which prints each token as it is being optimized.
- `CTX_OUTPUT.EVENT_INDEX_PRINT_TOKEN`, which prints the each token as it is being indexed.
- `CTX_OUTPUT.EVENT_DRG_DUMP_ERRORSTACK`, which prints the stack trace for the specified DRG error in the log. An error will be raised if `errnum` is not specified.

**errnum**
Specify the DRG error number for a `CTX_OUTPUT.EVENT_DRG_DUMP_ERRORSTACK` event.

Example

```sql
begin
  CTX_OUTPUT.ADD_EVENT_CTX_OUTPUT.EVENT_INDEX_PRINT_ROWID);
end;
```

Related Topics

See Also: "REMOVE_EVENT" on page 10-12
ADD_TRACE

Use this procedure to enable a trace. If the trace has not been enabled, this call adds the trace to the list of active traces and resets its value to 0. If the trace has already been enabled, an error is raised.

Syntax

```
CTX_OUTPUT.ADD_TRACE(trace_id BINARY_INTEGER);
```

**trace_id**
Specify the ID of the trace to enable. See Table 10–1 for possible trace values.

Notes

Table 10–1 shows the available traces:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>ID</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRACE_IDX_USER_DATASTORE</td>
<td>1</td>
<td>Time spent executing user datastore</td>
</tr>
<tr>
<td>TRACE_IDX_AUTO_FILTER</td>
<td>2</td>
<td>Time spent invoking the AUTO_FILTER filter. (Replaces the deprecated TRACE_IDX_INSO_FILTER trace)</td>
</tr>
<tr>
<td>TRACE_QRY_XX_TIME</td>
<td>3</td>
<td>Time spent executing the $X cursor</td>
</tr>
<tr>
<td>TRACE_QRY_XF_TIME</td>
<td>4</td>
<td>Time spent fetching from $X</td>
</tr>
<tr>
<td>TRACE_QRY_X_ROWS</td>
<td>5</td>
<td>Total number of rows whose token metadata was fetched from $X</td>
</tr>
<tr>
<td>TRACE_QRY_IF_TIME</td>
<td>6</td>
<td>Time spent fetching the LOB locator from $I</td>
</tr>
<tr>
<td>TRACE_QRY_IR_TIME</td>
<td>7</td>
<td>Time spent reading $I LOB information</td>
</tr>
<tr>
<td>TRACE_QRY_I_ROWS</td>
<td>8</td>
<td>Number of rows whose $I token_info was actually read</td>
</tr>
<tr>
<td>TRACE_QRY_I_SIZE</td>
<td>9</td>
<td>Number of bytes read from $I LOBs</td>
</tr>
<tr>
<td>TRACE_QRY_R_TIME</td>
<td>10</td>
<td>Time spent fetching and reading $R information</td>
</tr>
<tr>
<td>TRACE_QRY_CON_TIME</td>
<td>11</td>
<td>Time spent in CONTAINS processing (drexrcontains/drexrstart/drexrfetch)</td>
</tr>
<tr>
<td>TRACE_QRY_S_TIME</td>
<td>15</td>
<td>Time spent fetching and reading $S$ information</td>
</tr>
</tbody>
</table>

Tracing is independent of logging. Logging does not have to be on to start tracing, and vice-versa. Traces are associated with a session—they can measure operations that take place within a single session, and conversely, cannot make measurements across sessions. During parallel sync or optimize, the trace profile will be copied to the slave sessions if and only if tracing is currently enabled. Each slave will accumulate its own traces and implicitly write all trace values to the slave logfile before termination.
Related Topics

See Also:  "REMOVE_TRACE" on page 10-13, "GET_TRACE_VALUE" on page 10-9, "LOG_TRACES" on page 10-10, and "RESET_TRACE" on page 10-14, as well as Oracle Text Application Developer’s Guide
DISABLE_QUERY_STATS

Disables gathering of query stats for the index.

Syntax

cxt_output.disable_query_stats(
    index_name IN VARCHAR2
);

index_name
The name of the index on which query stats collection is to be disabled.

Example

Turn off gathering of query stats for the index myindex.

CTX_OUTPUT.DISABLE_QUERY_STATS(myindex);

Notes

Once the query stats is disabled for an index, gathering and storing query-related metadata is stopped for that index. All the entries corresponding to that index are cleared from the dictionary tables. An error is returned if you call this procedure on an index where query stats is not enabled.

Related Topics

CTX_OUTPUT.ENABLE_QUERY_STATS on page 10-6, and CTX_REPORT.INDEX_STATS on page 12-8.
ENABLE_QUERY_STATS

Enables gathering of query stats for the index. To have query-related metadata stored for the index, use this procedure to enable collection of statistics on that index. You can only access the gathered metadata when `ctx_output.enable_query_stats` is turned on for the index.

**Note:** Accessing the query stats metadata only works when `ctx_output.enable_query_stats` is turned on for the index. Please see `CTX_REPORT.INDEX_STATS` for the list of gathered query stats metadata.

**Syntax**

```sql
ctx_output.enable_query_stats(
  index_name IN VARCHAR2
);
```

**index_name**
The name of the index on which query stats collection is to be enabled.

**Example**

Turn on gathering of query stats for the index `myindex`.

```sql
CTX_OUTPUT.ENABLE_QUERY_STATS('myindex');
```

**Notes**

The information that shows whether query stats is enabled on an index is available in the views: `CTX_INDEXES` and `CTX_USER_INDEXES` under the column `idx_query_stats_enabled`, which is in both of these views. If `query_stats` is enabled for an index, then the column displays YES; if not, then the column displays NO.

The data corresponding to the query statistics will be stored in persistent dictionary tables. Once `statistics` has been enabled for a particular index, query statistics will be collected for that index from all sessions.

If you call this procedure for an index where query stats is already enabled, then an error is thrown.

Statistics collection has a minimal effect on query performance.

**Related Topics**

`CTX_OUTPUT.DISABLE_QUERY_STATS`, and `CTX_REPORT.INDEX_STATS`. 
This procedure halts logging index and document service requests.

**Syntax**

```plaintext
ctx_output.end_log;
```

**Example**

```plaintext
begin
  CTX_OUTPUT.END_LOG;
  end;
```
Use this procedure to stop logging queries into a logfile created with CTX_OUTPUT.START_QUERY_LOG.

Syntax

cxt_output.end_query_log;

Example

begin
  CTX_OUTPUT.START_QUERY_LOG('mylog1');
  < get queries >
  CTX_OUTPUT.END_QUERY_LOG;
end;
GET_TRACE_VALUE

Use this procedure to programmatically retrieve the current value of a trace.

Syntax

```
CTX_OUTPUT.GET_TRACE_VALUE(trace_id BINARY_INTEGER);
```

**trace_id**
Specify the trace ID whose value you want. See Table 10-1, "Available Traces" on page 10-3 for possible values.

Example

This sets the value of the variable `value`:

```
value := ctx_output.get_trace_value(trace_id);
```

Notes

You can also retrieve trace values through SQL:

```
select * from ctx_trace_values;
```

See "CTX_TRACE_VALUES" on page G-12 for the entries in the CTX_TRACE_VALUES view.

If the trace has not been enabled, an error is raised.

Traces are not reset to 0 by this call.

Traces are associated with a session—they can measure operations that take place within a single session, and conversely, cannot make measurements across sessions.

Related Topics

See Also:  "REMOVE_TRACE" on page 10-13, "ADD_TRACE" on page 10-3, "LOG_TRACES" on page 10-10,and "RESET TRACE" on page 10-14, as well as the Oracle Text Application Developer’s Guide
Use this procedure to print all active traces to the logfile.

Syntax

    CTX_OUTPUT.LOG_TRACES;

Notes

If logging has not been started, an error is raised.
Traces are not reset to 0 by this call.

This procedure looks for the logfile in the directory specified by the LOG_DIRECTORY system parameter, which is $ORACLE_HOME/ctx/log on UNIX. You can query the CTX_PARAMETERS view to find the current setting.

Related Topics

See Also:  "REMOVE_TRACE" on page 10-13, "GET_TRACE_VALUE" on page 10-9, "ADD_TRACE" on page 10-3, and "RESET_TRACE" on page 10-14, as well as the Oracle Text Application Developer’s Guide
LOGFILENAME

Returns the filename for the current log. This procedure looks for the logfile in the directory specified by the LOG_DIRECTORY system parameter, which is $ORACLE_HOME/ctx/log on UNIX. You can query the CTX_PARAMETERS view to find the current setting.

Syntax

```sql
CTX_OUTPUT.LOGFILENAME RETURN VARCHAR2;
```

Returns

Log file name.

Example

```sql
declare
    logname varchar2(100);
begin
    logname := CTX_OUTPUT.LOGFILENAME;
    dbms_output.put_line('The current log file is: '||logname);
end;
```
Use this procedure to remove an event from the index log.

Syntax

```
CTX_OUTPUT.REMOVE_EVENT(event in NUMBER);
```

**event**

Specify the type of index event to remove from the log. You can remove the following events:

- `CTX_OUTPUT.EVENT_INDEX_PRINT_ROWID`, which logs the rowid of each row after it is indexed. This is useful for debugging a failed index operation.
- `CTX_OUTPUT.EVENT_OPT_PRINT_TOKEN`, which prints each token as it is being optimized.
- `CTX_OUTPUT.EVENT_INDEX_PRINT_TOKEN`, which prints the each token as it is being indexed.

Example

```
begin
    CTX_OUTPUT.REMOVE_EVENT(CTX_OUTPUT.EVENT_INDEX_PRINT_ROWID);
end;
```

Related Topics

See Also:  
"ADD_EVENT" on page 10-2
REMOVE_TRACE

Use this procedure to disable a trace.

Syntax

\[
\text{CTX\_OUTPUT.REMOVE\_TRACE(trace\_id BINARY\_INTEGER)};
\]

**trace\_id**
Specify the ID of the trace to disable. See Table 10–1, "Available Traces" on page 10-3 for possible values.

Notes

If the trace has not been enabled, an error is raised.

Related Topics

See Also: "GET\_TRACE\_VALUE" on page 10-9, "ADD\_TRACE" on page 10-3, "LOG\_TRACES" on page 10-10, and "RESET\_TRACE" on page 10-14, as well as the Oracle Text Application Developer’s Guide
RESET_TRACE

Use this procedure to clear a trace (that is, reset it to 0).

Syntax

```sql
CTX_OUTPUT.RESET_TRACE(trace_id BINARY_INTEGER);
```

**trace_id**
Specify the ID of the trace to reset. See Table 10-1, "Available Traces" on page 10-3 for possible values.

Notes

If the trace has not been enabled, an error is raised.

Related Topics

See Also: "REMOVE_TRACE" on page 10-13, "GET_TRACE_VALUE" on page 10-9, "ADD_TRACE" on page 10-3, "LOG_TRACES" on page 10-10, as well as the Oracle Text Application Developer’s Guide
START_LOG

Begin logging index and document service requests.

Syntax

CTX_OUTPUT.START_LOG(logfile in varchar2, overwrite in default true);

logfile
Specify the name of the log file. The log is stored in the directory specified by the system parameter LOG_DIRECTORY.

overwrite
Specify whether you want to overwrite or append to the original query log file specified by logfile, if it already exists. The default is to overwrite the original query log file.

Example

begin
 CTX_OUTPUT.START_LOG('mylog1');
end;

Notes

Logging is independent of tracing. Logging does not have to be on to start tracing, and vice-versa.

Logging is associated with a session—it can log operations that take place within a single session, and, conversely, cannot make measurements across sessions.

Filenames used in CTX_OUTPUT.START_LOG are restricted to the following characters: alphanumeric, minus, period, space, hash, underscore, single and double quotes. Any other character in the filename will raise an error.
Begin logging query requests into a query log file.

Use CTX_OUTPUT.END_QUERY_LOG to stop logging queries. Use CTX_REPORT.QUERY_LOG_SUMMARY to obtain reports on logged queries, such as which queries returned successfully the most times.

The query log includes the query string, the index name, and the timestamp of the query, as well as whether or not the query successfully returned a hit. A successful query for the phrase *Blues Guitarists* made at 6:46 (local time) on November 11th, 2003, would be entered into the query log in this form:

```
<QuerySet><TimeStamp>18:46:51 02/04/03</TimeStamp><IndexName>IDX_SEARCH_TABLE</IndexName><Query>Blues Guitarists</Query><ReturnHit>Yes</ReturnHit></QuerySet>
```

**Syntax**

```sql
CTX_OUTPUT.START_QUERY_LOG(logfile in varchar2, overwrite in default true);
```

**logfile**
Specify the name of the query log file. The query log is stored in the directory specified by the system parameter LOG_DIRECTORY.

**overwrite**
Specify whether you want to overwrite or append to the original query log file specified by *logfile*, if it already exists. The default is to overwrite the original query log file.

**Example**

```sql
begin
CTX_OUTPUT.START_QUERY_LOG('mylog1');
< get queries >
CTX_OUTPUT.END_QUERY_LOG;
end;
```

**Notes**

Filenames used in CTX_OUTPUT.START_QUERY_LOG are restricted to the following characters: alphanumeric, minus, period, space, hash, underscore, single and double quotes. Any other character in the filename will raise an error.

Logging is associated with a session—it can log operations that take place within a single session, and, conversely, cannot make measurements across sessions.
This chapter describes the CTX_QUERY PL/SQL package you can use for generating query feedback, counting hits, and creating stored query expressions.

**Note:** You can use this package only when your index type is CONTEXT. This package does not support the CTXCAT index type.

The CTX_QUERY package includes the following procedures and functions:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BROWSE_WORDS</td>
<td>Returns the words around a seed word in the index.</td>
</tr>
<tr>
<td>COUNT_HITS</td>
<td>Returns the number hits to a query.</td>
</tr>
<tr>
<td>EXPLAIN</td>
<td>Generates query expression parse and expansion information.</td>
</tr>
<tr>
<td>HFEEDBACK</td>
<td>Generates hierarchical query feedback information (broader term, narrower term, and related term).</td>
</tr>
<tr>
<td>REMOVE_SQE</td>
<td>Removes a specified stored query expression from the SQL tables.</td>
</tr>
<tr>
<td>RESULT_SET</td>
<td>Executes a query and generates a result set.</td>
</tr>
<tr>
<td>STORE_SQE</td>
<td>Executes a query and stores the results in stored query expression tables.</td>
</tr>
</tbody>
</table>
This procedure enables you to browse words in an Oracle Text index. Specify a seed word and `BROWSE_WORDS` returns the words around it in the index, and an approximate count of the number of documents that contain each word.

This feature is useful for refining queries. You can identify the following words:

- Unselective words (words that have low document count)
- Misspelled words in the document set

Syntax 1: To Store Results in Table

```sql
ctx_query.browse_words(
    index_name  IN   VARCHAR2,
    seed        IN   VARCHAR2,
    restab      IN   VARCHAR2,
    browse_id   IN   NUMBER   DEFAULT 0,
    numwords    IN   NUMBER   DEFAULT 10,
    direction   IN   VARCHAR2 DEFAULT BROWSE_AROUND,
    part_name   IN   VARCHAR2 DEFAULT NULL
);
```

Syntax 2: To Store Results in Memory

```sql
ctx_query.browse_words(
    index_name  IN      VARCHAR2,
    seed        IN      VARCHAR2,
    resarr      IN OUT  BROWSE_TAB,
    numwords    IN      NUMBER   DEFAULT 10,
    direction   IN      VARCHAR2 DEFAULT BROWSE_AROUND,
    part_name   IN      VARCHAR2 DEFAULT NULL
);
```

**index**

Specify the name of the index. You can specify `schema.name`. Must be a local index.

**seed**

Specify the seed word. This word is lexed before browse expansion. The word need not exist in the token table. `seed` must be a single word. Using multiple words as the seed will result in an error.

**restab**

Specify the name of the result table. You can enter `restab` as `schema.name`. The table must exist before you call this procedure, and you must have `INSERT` permissions on the table. This table must have the following schema.

<table>
<thead>
<tr>
<th>Column</th>
<th>Datatype</th>
</tr>
</thead>
<tbody>
<tr>
<td>browse_id</td>
<td>number</td>
</tr>
<tr>
<td>word</td>
<td>varchar2(64)</td>
</tr>
<tr>
<td>doc_count</td>
<td>number</td>
</tr>
</tbody>
</table>

Existing rows in `restab` are not deleted before `BROWSE_WORDS` is called.
resarr
Specify the name of the result array. resarr is of type ctx_query.browse_tab.

type browse_rec is record {
    word varchar2(64),
    doc_count number
};
type browse_tab is table of browse_rec index by binary_integer;

browse_id
Specify a numeric identifier between 0 and $2^{32}$. The rows produced for this browse have a value of in the browse_id column in restab. When you do not specify browse_id, the default is 0.

numwords
Specify the number of words returned.

direction
Specify the direction for the browse. You can specify one of:

<table>
<thead>
<tr>
<th>value</th>
<th>behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEFORE</td>
<td>Browse seed word and words alphabetically before the seed.</td>
</tr>
<tr>
<td>AROUND</td>
<td>Browse seed word and words alphabetically before and after the seed.</td>
</tr>
<tr>
<td>AFTER</td>
<td>Browse seed word and words alphabetically after the seed.</td>
</tr>
</tbody>
</table>

Symbols CTX_QUERY.BROWSE_BEFORE, CTX_QUERY.BROWSE_AROUND, and CTX_QUERY.BROWSE_AFTER are defined for these literal values as well.

part_name
Specify the name of the index partition to browse.

Example

Browsing Words with Result Table
begin
ctx_query.browse_words('myindex','dog','myres',numwords=>5,direction=>'AROUND');
end;

select word, doc_count from myres order by word;

<table>
<thead>
<tr>
<th>WORD</th>
<th>DOC_COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>CZAR</td>
<td>15</td>
</tr>
<tr>
<td>DARLING</td>
<td>5</td>
</tr>
<tr>
<td>DOC</td>
<td>73</td>
</tr>
<tr>
<td>DUNK</td>
<td>100</td>
</tr>
<tr>
<td>EAR</td>
<td>3</td>
</tr>
</tbody>
</table>

Browsing Words with Result Array
set serveroutput on;
declare
  resarr ctx_query.browse_tab;
begin
ctx_query.browse_words('myindex','dog',resarr,5,CTX_QUERY.BROWSE_AROUND);
for i in 1..resarr.count loop
dbms_output.put_line(resarr(i).word || ' : ' || resarr(i).doc_count);
end loop;
end;
COUNT_HITS

Returns the number of hits for the specified query. You can call COUNT_HITS in exact or estimate mode. Exact mode returns the exact number of hits for the query. Estimate mode returns an upper-bound estimate but runs faster than exact mode.

Syntax

Syntax 1
exec CTX_QUERY.COUNT_HITS(
    index_name  IN VARCHAR2,
    text_query  IN VARCHAR2,
    exact       IN BOOLEAN  DEFAULT TRUE,
    part_name   IN VARCHAR2 DEFAULT NULL
) RETURN NUMBER;

Syntax 2
exec CTX_QUERY.COUNT_HITS_CLOB_QUERY(
    index_name  IN VARCHAR2,
    text_query  IN CLOB,
    exact       IN BOOLEAN DEFAULT TRUE,
    part_name   IN VARCHAR2 DEFAULT NULL
) RETURN NUMBER;

index_name
Specify the index name.

text_query
Specify the query.

exact
Specify TRUE for an exact count. Specify FALSE for an upper-bound estimate.

Specifying FALSE returns a less accurate number but runs faster. Specifying FALSE might return a number which is too high if rows have been updated or deleted since the last FULL index optimize. Optimizing in full mode removes these false hits, and then EXACT set to FALSE will return the same number as EXACT set to TRUE.

part_name
Specify the name of the index partition to query.

Notes

If the query contains structured criteria, then you should use SELECT COUNT(*)

If the index was created with the TRANSACTIONAL parameter, then COUNT_HITS will include pending rowids as well as those that have been synchronized.
Use CTX_QUERY.EXPLAIN to generate explain plan information for a query expression. The EXPLAIN plan provides a graphical representation of the parse tree for a Text query expression. This information is stored in a result table.

This procedure does not execute the query. Instead, this procedure can tell you how a query is expanded and parsed before you enter the query. This is especially useful for stem, wildcard, thesaurus, fuzzy, soundex, or about queries. Parse trees also show the following information:

- Order of execution (precedence of operators)
- ABOUT query normalization
- Query expression optimization
- Stop-word transformations
- Breakdown of composite-word tokens

Knowing how Oracle Text evaluates a query is useful for refining and debugging queries. You can also design your application so that it uses the explain plan information to help users write better queries.

**Syntax**

**Syntax 1**
```
exec CTX_QUERY.EXPLAIN(
  index_name     IN VARCHAR2,
  text_query     IN VARCHAR2,
  explain_table  IN VARCHAR2,
  sharelevel     IN NUMBER DEFAULT 0,
  explain_id     IN VARCHAR2 DEFAULT NULL,
  part_name      IN VARCHAR2 DEFAULT NULL
);
```

**Syntax 2**
```
exec CTX_QUERY.EXPLAIN_CLOB_QUERY(
  index_name     IN VARCHAR2,
  text_query     IN CLOB,
  explain_table  IN VARCHAR2,
  sharelevel     IN NUMBER DEFAULT 0,
  explain_id     IN VARCHAR2 DEFAULT NULL,
  part_name      IN VARCHAR2 DEFAULT NULL
);
```

**index_name**
Specify the name of the index to be queried.

**text_query**
Specify the query expression to be used as criteria for selecting rows.

When you include a wildcard, fuzzy, or soundex operator in text_query, this procedure looks at the index tables to determine the expansion.

Wildcard, fuzzy (?), and soundex (!) expression feedback does not account for lazy deletes as in regular queries.
**explain_table**
Specify the name of the table used to store representation of the parse tree for *text_query*. You must have at least INSERT and DELETE privileges on the table used to store the results from EXPLAIN.

**See Also:** "EXPLAIN Table" in Appendix A, "Oracle Text Result Tables" for more information about the structure of the explain table.

**sharelevel**
Specify whether explain_table is shared by multiple EXPLAIN calls. Specify 0 for exclusive use and 1 for shared use. Default is 0 (single-use).

When you specify 0, the system automatically truncates the result table before the next call to EXPLAIN.

When you specify 1 for shared use, this procedure does not truncate the result table. Only results with the same explain_id are updated. When no results with the same explain_id exist, new results are added to the EXPLAIN table.

**explain_id**
Specify a name that identifies the explain results returned by an EXPLAIN procedure when more than one EXPLAIN call uses the same shared EXPLAIN table. Default is NULL.

**part_name**
Specify the name of the index partition to query.

**Example**

**Creating the Explain Table**
To create an explain table called test_explain for example, use the following SQL statement:

```sql
create table test_explain(
    explain_id varchar2(30),
    id number,
    parent_id number,
    operation varchar2(30),
    options varchar2(30),
    object_name varchar2(64),
    position number,
    cardinality number);
```

**Running CTX_QUERY.EXPLAIN**
To obtain the expansion of a query expression such as `comp% OR ?smith`, use CTX_QUERY.EXPLAIN as follows:

```sql
ctx_query.explain{
    index_name => 'newindex',
    text_query => 'comp% OR ?smith',
    explain_table => 'test_explain',
    sharelevel => 0,
    explain_id => 'Test'};
```

**Retrieving Data from Explain Table**
To read the explain table, you can select the columns as follows:
select explain_id, id, parent_id, operation, options, object_name, position
from test_explain order by id;

The output is ordered by ID to simulate a hierarchical query:

<table>
<thead>
<tr>
<th>EXPLAIN_ID</th>
<th>ID</th>
<th>PARENT_ID</th>
<th>OPERATION</th>
<th>OPTIONS</th>
<th>OBJECT_NAME</th>
<th>POSITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>1</td>
<td>0 OR</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>1</td>
</tr>
<tr>
<td>Test</td>
<td>2</td>
<td>1 EQUIVALENCE</td>
<td>NULL</td>
<td>COMP%</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>3</td>
<td>2 WORD</td>
<td>NULL</td>
<td>COMPTROLLER</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>4</td>
<td>2 WORD</td>
<td>NULL</td>
<td>COMPUTER</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>5</td>
<td>1 EQUIVALENCE</td>
<td>(?)</td>
<td>SMITH</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>6</td>
<td>5 WORD</td>
<td>NULL</td>
<td>SMITH</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>7</td>
<td>5 WORD</td>
<td>NULL</td>
<td>SMYTHE</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Restrictions

CTX_QUERY.EXPLAIN does not support the use of query templates.

You cannot use CTX_QUERY.EXPLAIN with remote queries.

If the query utilizes themes (for example, with an ABOUT query), then a knowledge base must be installed. Such a knowledge base may or may not have been installed with Oracle Text. For more information on knowledge bases, see Oracle Text Application Developer's Guide.

Related Topics

Chapter 3, "Oracle Text CONTAINS Query Operators"

Appendix H, "Stopword Transformations in Oracle Text"
In English or French, this procedure generates hierarchical query feedback information (broader term, narrower term, and related term) for the specified query. Broader term, narrower term, and related term information is obtained from the knowledge base. However, only knowledge base terms that are also in the index are returned as query feedback information. This increases the chances that terms returned from HFEEDBACK produce hits over the currently indexed document set.

Hierarchical query feedback information is useful for suggesting other query terms to the user.

**Note:** CTX_QUERY.HFEEDBACK requires an installed knowledge base. A knowledge base may or may not have been installed with Oracle Text. For more information on knowledge bases, see Oracle Text Application Developer's Guide.

CTX_QUERY.HFEEDBACK is only supported in English and French.

### Syntax

**Syntax 1**

```sql
exec CTX_QUERY.HFEEDBACK{
    index_name     IN VARCHAR2,
    text_query     IN VARCHAR2,
    feedback_table IN VARCHAR2,
    sharelevel     IN NUMBER DEFAULT 0,
    feedback_id    IN VARCHAR2 DEFAULT NULL,
    part_name      IN VARCHAR2 DEFAULT NULL
};
```

**Syntax 2**

```sql
exec CTX_QUERY.HFEEDBACK_CLOB_QUERY{
    index_name     IN VARCHAR2,
    text_query     IN CLOB,
    feedback_table IN VARCHAR2,
    sharelevel     IN NUMBER DEFAULT 0,
    feedback_id    IN VARCHAR2 DEFAULT NULL,
    part_name      IN VARCHAR2 DEFAULT NULL
};
```

**index_name**

Specify the name of the index for the text column to be queried.

**text_query**

Specify the query expression to be used as criteria for selecting rows.

**feedback_table**

Specify the name of the table used to store the feedback terms.

**See Also:** "HFEEDBACK Table" in Appendix A, "Oracle Text Result Tables" for more information about the structure of the explain table.
sharelevel
Specify whether feedback_table is shared by multiple HFEEDBACK calls. Specify 0 for exclusive use and 1 for shared use. Default is 0 (single-use).

When you specify 0, the system automatically truncates the feedback table before the next call to HFEEDBACK.

When you specify 1 for shared use, this procedure does not truncate the feedback table. Only results with the same feedback_id are updated. When no results with the same feedback_id exist, new results are added to the feedback table.

feedback_id
Specify a value that identifies the feedback results returned by a call to HFEEDBACK when more than one HFEEDBACK call uses the same shared feedback table. Default is NULL.

part_name
Specify the name of the index partition to query.

Example

Create HFEEDBACK Result Table
Create a result table to use with CTX_QUERY.HFEEDBACK as follows:

```sql
CREATE TABLE restab (  
  feedback_id VARCHAR2(30),  
  id NUMBER,  
  parent_id NUMBER,  
  operation VARCHAR2(30),  
  options VARCHAR2(30),  
  object_name VARCHAR2(80),  
  position NUMBER,  
  bt_feedback ctxsys.ctx_feedback_type,  
  rt_feedback ctxsys.ctx_feedback_type,  
  nt_feedback ctxsys.ctx_feedback_type  
) NESTED TABLE bt_feedback STORE AS res_bt  
NESTED TABLE rt_feedback STORE AS res_rt  
NESTED TABLE nt_feedback STORE AS res_nt;
```

CTX_FEEDBACK_TYPE is a system-defined type in the CTXSYS schema.

See Also: "HFEEDBACK Table" in Appendix A, "Oracle Text Result Tables" for more information about the structure of the HFEEDBACK table.

Call CTX_QUERY.HFEEDBACK
The following code calls the HFEEDBACK procedure with the query computer industry.

```sql
BEGIN  
  ctx_query.hfeedback (index_name => 'my_index',  
    text_query => 'computer industry',  
    feedback_table => 'restab',  
    sharelevel => 0,  
    feedback_id => 'query10'  
  );
END;
```
Select From the Result Table

The following code extracts the feedback data from the result table. It extracts broader term, narrower term, and related term feedback separately from the nested tables.

```
DECLARE
    i NUMBER;
BEGIN
    FOR frec IN (
        SELECT object_name, bt_feedback, rt_feedback, nt_feedback
        FROM restab
        WHERE feedback_id = 'query10' AND object_name IS NOT NULL
    ) LOOP
        dbms_output.put_line('Broader term feedback for ' || frec.object_name || ':
            ');
        i := frec.bt_feedback.FIRST;
        WHILE i IS NOT NULL LOOP
            dbms_output.put_line(frec.bt_feedback(i).text);
            i := frec.bt_feedback.NEXT(i);
        END LOOP;
        dbms_output.put_line('Related term feedback for ' || frec.object_name || ':
            ');
        i := frec.rt_feedback.FIRST;
        WHILE i IS NOT NULL LOOP
            dbms_output.put_line(frec.rt_feedback(i).text);
            i := frec.rt_feedback.NEXT(i);
        END LOOP;
        dbms_output.put_line('Narrower term feedback for ' || frec.object_name || ':
            ');
        i := frec.nt_feedback.FIRST;
        WHILE i IS NOT NULL LOOP
            dbms_output.put_line(frec.nt_feedback(i).text);
            i := frec.nt_feedback.NEXT(i);
        END LOOP;
    END LOOP;
END;
```

Sample Output

The following output is for the preceding example, which queries on `computer industry`:

Broader term feedback for computer industry:
- hard sciences
- Related term feedback for computer industry:
  - computer networking
  - electronics
  - knowledge
  - library science
  - mathematics
  - optical technology
  - robotics
  - satellite technology
  - semiconductors and superconductors
  - symbolic logic
  - telecommunications industry
Narrower term feedback for computer industry:
- ABEND - abnormal end of task
- AT&T Starlans
ATI Technologies, Incorporated
ActivCard
Actrade International Ltd.
Alta Technology
Amiga Format
Amiga Library Services
Amiga Shopper
Amstrat Action
Apple Computer, Incorporated
...

---

**Note:** The HFEEDBACK information you obtain depends on the contents of your index and knowledge base and as such might differ from the sample shown.

---

**Restrictions**

CTX_QUERY.HFEEDBACK does not support the use of query templates.
The `CTX_QUERY.REMOVE_SQE` procedure removes the specified stored query expression.

**Syntax**

```sql
CTX_QUERY.REMOVE_SQE(query_name IN VARCHAR2);
```

**query_name**
Specify the name of the stored query expression to be removed.

**Example**

```sql
begin
    ctx_query.remove_sqe('disasters');
end;
```
This procedure executes an XML query and generates a result set in XML. The result set interface can return data views that are difficult to express in SQL, such as top N by category queries.

**See Also:** Oracle Text Application Developer’s Guide for details on how to use the result set interface

### Syntax

```
CTX_QUERY.RESULT_SET (
    index_name            IN VARCHAR2,
    query                 IN VARCHAR2,
    result_set_descriptor IN CLOB,
    result_set            IN OUT NOCOPY CLOB,
    part_name             IN VARCHAR2 DEFAULT NULL);
```

- **index_name**
  Specify the index against which to execute the query.

- **query**
  Specify the query string.

- **result_set_descriptor**
  Specify the result set descriptor in XML. It describes what the result set should contain. See “The Input Result Set Descriptor” on page 11-14 for more details.

- **result_set**
  Specify the output result set. If this variable is NULL on input, a session-duration temporary lob will be allocated and returned to the user. The user is responsible for deallocating this temporary lob. See “The Output Result Set XML” on page 11-16 for more details.

- **part_name**
  Specify the index partition name. If the index is global, part_name must be NULL. If the index is partitioned and part_name is not NULL, then the query will only be evaluated for the given partition. If the index is partitioned and part_name is NULL, then the query will be evaluated for all partitions.

### The Input Result Set Descriptor

The result set descriptor is an XML message which describes what to calculate for the result set. The elements present in the result set descriptor and the order in which they occur serve as a simple template, specifying what to include in the output result set. That is, there should be the list of hit rowids, then a count, then a token count, and so on. The attributes of the elements specify the parameters and options to the specific operations, such as number of hits in the list of rowids, estimate versus exact count, and so on.

The result set descriptor itself is XML conforming to the following DTD:

```xml
<!ELEMENT ctx_result_set_descriptor (hitlist?, group*, count?)>
<!ELEMENT hitlist (rowid?, score?, sdata*)>
<!ELEMENT group(count?)>
<!ELEMENT count EMPTY>
<!ELEMENT rowid EMPTY>
```
The following is a description of the possible XML elements for the result set descriptor:

- **ctx_result_set_descriptor**
  
  This is the root element for the result set descriptor. The parent element is none, as are the available attributes.
  
  The possible child elements are:
  - Zero or more hitlist elements.
  - Zero or more group elements.
  - At most one count element.

- **group**

  The group element causes the generated result set to include a group breakdown. In other words, a breakdown of the results by SDATA section values. The parent element is ctx_result_set_descriptor, and the available attributes are:
  
  - sdata
    
    Specifies the name of the SDATA section to use for grouping. It is required.
  
  Possible child elements of group are:
  - At most one count element.

- **hitlist**

  The hitlist element controls inclusion of a list of hit documents. The parent element is ctx_result_set_descriptor, and the available attributes are:
  
  - start_hit_num
    
    This specifies the starting document hit to be included in the generated result set. This can be set to any positive integer less than or equal to 2048. For example, if start_hit_num is 21, then the result set will include document hits starting from the 21st document hit. This element is required.
  
  - end_hit_num
    
    This specifies the last document hit to be included in the generated result set. This can be set to any positive integer less than or equal to 2048. For example, if end_hit_num is 40, then the result set will include document hits up to the 40th document hit. This element is required.
  
  - order
    
    This is an optional attribute that specifies the order for the documents in the generated result set. The value is a list similar to a SQL ORDER BY statement, except that, instead of column names, they can either be SCORE or SDATA section names. In the following example, MYDATE and MYPRICE are the SDATA section names:
    
    \( \text{order = "SCORE DESC, MYDATE, MYPRICE DESC"} \)
The possible child elements are:
- At most one rowid element.
- At most one score element.
- At most one sdata element.

■ count
This element causes the generated result set to include a count of the number of hit documents. The parent elements are:
- ctx_result_set_descriptor
- group
The available attributes are:
- exact
  This is to estimate mode. Set to true or false. It is required, and the default is false.
The possible child elements are none.

■ rowid
This child element causes the generated result set to include rowid information for each hit. The parent element is hitlist. There are no attributes and no possible child elements.

■ score
This child element causes the generated result set to include score information for each hit.
- The parent element is hitlist.
- There are no available attributes, and no possible child elements.

■ sdata
This child element causes the generated result set to include sdata values for each hit.
- The parent element is hitlist.
- The available attribute is name. This specifies the name of the sdata section. It is required.
- There are no child elements.

The Output Result Set XML
The output result set XML is XML conforming to the following DTD:

```xml
<!ELEMENT ctx_result_set (hitlist?, groups*, count?)>
<!ELEMENT hitlist (hit*)>
<!ELEMENT hit(rowid?, score?, sdata*)>
<!ELEMENT groups (group*)>
<!ELEMENT group (count?)>
<!ELEMENT count CDATA>
<!ELEMENT rowid CDATA>
<!ELEMENT score CDATA>
<!ELEMENT sdata CDATA>
<!ATTLIST groups sdata CDATA #REQUIRED>
```
The following is a description of the list of possible XML elements for the output result set:

- **ctx_result_set**
  - This is the root element for the generated result set. There are no attributes. The parent is none. The possible child elements are:
    - At most one hitlist element.
    - Zero or more groups elements.

- **groups**
  - This delimits the start of a group breakdown section. The parent element is ctx_result_set. The available attributes are:
    - sdata
      - This is the name of the sdata section used for grouping.
    - The possible child elements are:
      - Zero or more group elements.

- **group**
  - This delimits the start of a GROUP BY value. The parent element is the groups element. The available attributes are:
    - value
      - This is the value of the sdata section.
    - The possible child elements are at most one count element.

- **hitlist**
  - This delimits the start of hitlist information. The parent element is ctx_result_set, while the children are zero or more hit elements. There are no attributes.

- **hit**
  - This delimits the start of the information for a particular document within a hitlist. The parent element is hitlist, and there are no available attributes. The possible child elements are:
    - Zero or one rowid elements.
    - Zero or one score element.
    - Zero or one sdata element.

- **rowid**
  - This is the rowid of the document, so the content is the rowid of the document. The parent element is the hit element. There are no child elements, and no available attributes.

- **score**
  - This is the score of the document. The parent element is the hit element. The content is the numeric score. There are no available attributes, and no possible child elements.
sdata
This is the SDATA value or values for the document. The parent element is the hit element, and the available attribute is name, which is the name of the sdata section. There are no possible child elements available. The content is the SDATA section value, which, for DATE values, is in the format "YYYY-MM-DD HH24:MI:SS", depending upon the actual values being stored.

count
This is the document hit count. The parent element is the ctx_result_set element or the group element. It contains the numeric hit count, and has no attributes, and no possible child elements.

Example
This call to ctx_query.result_set() with the specified XML result_set_descriptor will generate the following information in the form of XML:

- top 20 hits displaying, score, rowid, author SDATA section value, and pubDate SDATA section value, order by pubDate SDATA section value DESC and score DESC
- total doc hit count for the text query
- counts group by pubDate SDATA section values
- counts group by author SDATA section values

declare
rs clob;
begin
    dbms_lob.createtemporary(rs, true, dbms_lob.session);
    ctx_query.result_set('docidx', 'oracle', '
    <ctx_result_set_descriptor>
    <count/>
    <hitlist start_hit_num="1" end_hit_num="5" order="pubDate desc, score desc">
    <score/>
    <rowid/>
    <sdata name="author"/>
    <sdata name="pubDate"/>
    </hitlist>
    <group sdata="pubDate">
    <count/>
    </group>
    </ctx_result_set_descriptor>
    ', rs);
    dbms_lob.freetemporary(rs);
exception
    when others then
    dbms_lob.freetemporary(rs);
    raise;
end;
/

The XML output store in the result set output clob will resemble the following:

<ctx_result_set>
<hitlist>
    <hit>
        <score>3</score><rowid>AAAPoEAABAAAMnA4AC</rowid>
<sdata name="AUTHOR">John</sdata>
<sdata name="PUBDATE">2001-01-03 00:00:00</sdata>
</hit>
<hit>
<score>3</score><rowid>AAAPoEAABAAAMWsAAG</rowid>
<sdata name="AUTHOR">John</sdata>
<sdata name="PUBDATE">2001-01-03 00:00:00</sdata>
</hit>
<hit>
<score>3</score><rowid>AAAPoEAABAAAMWsAAK</rowid>
<sdata name="AUTHOR">John</sdata>
<sdata name="PUBDATE">2001-01-03 00:00:00</sdata>
</hit>
<hit>
<score>3</score><rowid>AAAPoEAABAAAMWsAAO</rowid>
<sdata name="AUTHOR">John</sdata>
<sdata name="PUBDATE">2001-01-03 00:00:00</sdata>
</hit>
<hit>
<score>3</score><rowid>AAAPoEAABAAAMWsAAS</rowid>
<sdata name="AUTHOR">John</sdata>
<sdata name="PUBDATE">2001-01-03 00:00:00</sdata>
</hit>
</hitlist>
<count>100</count>
</ctx_result_set>
STORE_SQE

This procedure creates a stored query expression. Only the query definition is stored.

Supported Operators

Stored query expressions support all of the CONTAINS query operators. Stored query expressions also support all of the special characters and other components that can be used in a query expression, including other stored query expressions.

Privileges

Users are allowed to create and remove stored query expressions owned by them. Users are allowed to use stored query expressions owned by anyone. The CTXSYS user can create or remove stored query expressions for any user.

Syntax

Syntax 1

CTX_QUERY.STORE_SQE(query_name IN VARCHAR2,
                     text_query IN VARCHAR2);

Syntax 2

CTX_QUERY.STORE_SQE_CLOB_QUERY(query_name IN VARCHAR2,
                                 text_query IN CLOB);

query_name

Specify the name of the stored query expression to be created.

text_query

Specify the query expression to be associated with query_name.

Example

begin
  ctx_query.store_sqe('disasters', 'hurricanes | earthquakes');
end;
This chapter describes how to use the CTX_REPORT package to create reports on indexing and querying. These reports can help you troubleshoot problems or fine-tune your applications.

This chapter contains the following topics:

- Procedures in CTX_REPORT
- Using the Function Versions

See Also: Oracle Text Application Developer’s Guide for an overview of the CTX_REPORT package and how you can use the various procedures described in this chapter

## Procedures in CTX_REPORT

The CTX_REPORT package contains the following procedures:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESCRIBE_INDEX</td>
<td>Creates a report describing the index.</td>
</tr>
<tr>
<td>DESCRIBE_POLICY</td>
<td>Creates a report describing a policy.</td>
</tr>
<tr>
<td>CREATE_INDEX_SCRIPT</td>
<td>Creates a SQL*Plus script to duplicate the named index.</td>
</tr>
<tr>
<td>CREATE_POLICY_SCRIPT</td>
<td>Creates a SQL*Plus script to duplicate the named policy.</td>
</tr>
<tr>
<td>INDEX_SIZE</td>
<td>Creates a report to show the internal objects of an index, their tablespaces and used sizes.</td>
</tr>
<tr>
<td>INDEX_STATS</td>
<td>Creates a report to show the various statistics of an index.</td>
</tr>
<tr>
<td>QUERY_LOG_SUMMARY</td>
<td>Creates a report showing query statistics</td>
</tr>
<tr>
<td>TOKEN_INFO</td>
<td>Creates a report showing the information for a token, decoded.</td>
</tr>
<tr>
<td>TOKEN_TYPE</td>
<td>Translates a name and returns a numeric token type.</td>
</tr>
</tbody>
</table>

## Using the Function Versions

Some of the procedures in the CTX_REPORT package have function versions. You can call these functions as follows:

```sql
select ctx_report.describe_index('MYINDEX') from dual;
```

In SQL*Plus, to generate an output file to send to support, you can do:
```
set long 64000
set pages 0
set heading off
set feedback off
spool outputfile
select ctx_report.describe_index('MYINDEX') from dual;
spool off
```
CREATE INDEX

Creates a report describing the index. This includes the settings of the index metadata, the indexing objects used, the settings of the attributes of the objects, and index partition descriptions, if any.

You can call this operation as a procedure with an IN OUT CLOB parameter or as a function that returns the report as a CLOB.

Syntax

```sql
procedure CTX_REPORT.DESCRIBE_INDEX(
    index_name     IN VARCHAR2,
    report         IN OUT NOCOPY CLOB,
    report_format  IN VARCHAR2 DEFAULT FMT_TEXT
);

function CTX_REPORT.DESCRIBE_INDEX(
    index_name     IN VARCHAR2,
    report_format  IN VARCHAR2 DEFAULT FMT_TEXT
) return CLOB;
```

**index_name**
Specify the name of the index to describe.

**report**
Specify the CLOB locator to which to write the report.

If `report` is NULL, a session-duration temporary CLOB will be created and returned. It is the caller's responsibility to free this temporary CLOB as needed.

The `report` CLOB will be truncated before report is generated, so any existing contents will be overwritten by this call.

**report_format**
Specify whether the report should be generated as 'TEXT' or as 'XML'. TEXT is the default. You can also specify the values CTX_REPORT.FMT_TEXT or CTX_REPORT.FMT_XML.

Notes

CTX_REPORT.DESCRIBE_INDEX outputs FILTER BY and ORDER BY column information if the index is created with FILTER BY and/or ORDER BY clauses.

Related Topics

"CREATE INDEX" on page 1-36, and "ADD_SDATA_COLUMN" on page 7-14
DESCRIBE_POLICY

Creates a report describing the policy. This includes the settings of the policy metadata, the indexing objects used, the settings of the attributes of the objects.

You can call this operation as a procedure with an \texttt{IN OUT CLOB} parameter or as a function that returns the report as a \texttt{CLOB}.

**Syntax**

```sql
procedure CTX_REPORT.DESCRIBE_POLICY(
    policy_name    IN VARCHAR2,
    report         IN OUT NOCOPY CLOB,
    report_format  IN VARCHAR2 DEFAULT FMT_TEXT
);

function CTX_REPORT.DESCRIBE_POLICY(
    policy_name    IN VARCHAR2,
    report_format  IN VARCHAR2 DEFAULT FMT_TEXT
) return CLOB;
```

**report**

Specify the \texttt{CLOB} locator to which to write the report.

If `report` is NULL, a session-duration temporary \texttt{CLOB} will be created and returned. It is the caller's responsibility to free this temporary \texttt{CLOB} as needed.

The `report` \texttt{CLOB} will be truncated before `report` is generated, so any existing contents will be overwritten by this call.

**report_format**

Specify whether the report should be generated as 'TEXT' or as 'XML'. TEXT is the default. You can also specify the values \texttt{CTX_REPORT.FMT_TEXT} or \texttt{CTX_REPORT.FMT_XML}.

**policy_name**

Specify the name of the policy to describe.
CREATE_INDEX_SCRIPT

Creates a SQL*Plus script which will create a text index that duplicates the named text index.

The created script will include creation of preferences identical to those used in the named text index. However, the names of the preferences will be different.

You can call this operation as a procedure with an IN OUT CLOB parameter or as a function that returns the report as a CLOB.

Syntax

```sql
procedure CTX_REPORT.CREATE_INDEX_SCRIPT(
    index_name      in varchar2,
    report          in out nocopy clob,
    prefname_prefix in varchar2 default null
);

function CTX_REPORT.CREATE_INDEX_SCRIPT(
    index_name      in varchar2,
    prefname_prefix in varchar2 default null
) return clob;
```

**index_name**
Specify the name of the index.

**report**
Specify the CLOB locator to which to write the script.

If `report` is NULL, a session-duration temporary CLOB will be created and returned. It is the caller's responsibility to free this temporary CLOB as needed.

The `report` CLOB will be truncated before report is generated, so any existing contents will be overwritten by this call.

**prefname_prefix**
Specify optional prefix to use for preference names.

If `prefname_prefix` is omitted or NULL, `index_name` will be used. The `prefname_prefix` follows index length restrictions.

Notes

CTX_REPORT.CREATE_INDEX_SCRIPT will also generate necessary FILTER BY and ORDER BY clauses for CREATE INDEX statements.

Related Topics

"CREATE INDEX" on page 1-36.
CREATE_POLICY_SCRIPT

Creates a SQL*Plus script which will create a text policy that duplicates the named text policy.
The created script will include creation of preferences identical to those used in the named text policy.
You can call this operation as a procedure with an IN OUT CLOB parameter or as a function that returns the report as a CLOB.

Syntax

procedure CTX_REPORT.CREATE_POLICY_SCRIPT(
  policy_name      in varchar2,
  report           in out nocopy clob,
  prefname_prefix  in varchar2 default null
);

function CTX_REPORT.CREATE_POLICY_SCRIPT(
  policy_name      in varchar2,
  prefname_prefix  in varchar2 default null
) return clob;

policy_name
Specify the name of the policy.

report
Specify the locator to which to write the script.
If report is NULL, a session-duration temporary CLOB will be created and returned. It is the caller’s responsibility to free this temporary CLOB as needed.
The report CLOB will be truncated before report is generated, so any existing contents will be overwritten by this call.

prefname_prefix
Specify the optional prefix to use for preference names. If prefname_prefix is omitted or NULL, policy name will be used. prefname_prefix follows policy length restrictions.
INDEX_SIZE

Creates a report showing the internal objects of the text index or text index partition, and their tablespaces, allocated, and used sizes.

You can call this operation as a procedure with an IN OUT CLOB parameter, or as a function that returns the report as a CLOB.

Syntax

procedure CTX_REPORT.INDEX_SIZE(
   index_name     IN VARCHAR2,
   report         IN OUT NOCOPY CLOB,
   part_name      IN VARCHAR2 DEFAULT NULL,
   report_format  IN VARCHAR2 DEFAULT FMT_TEXT
);

function CTX_REPORT.INDEX_SIZE(
   index_name     IN VARCHAR2,
   part_name      IN VARCHAR2 DEFAULT NULL,
   report_format  IN VARCHAR2 DEFAULT FMT_TEXT
) return clob;

index_name
Specify the name of the index to describe

report
Specify the CLOB locator to which to write the report.

If report is NULL, a session-duration temporary CLOB will be created and returned. It is the caller’s responsibility to free this temporary CLOB as needed.

The report CLOB will be truncated before report is generated, so any existing contents will be overwritten by this call

part_name
Specify the name of the index partition (optional). If part_name is NULL, and the index is a local partitioned text index, then all objects of all partitions will be displayed. If part_name is provided, then only the objects of a particular partition will be displayed.

report_format
Specify whether the report should be generated as 'TEXT' or as 'XML'. TEXT is the default. You can also specify the values CTX_REPORT.FMT_TEXT or CTX_REPORT.FMT_XML.

Notes

CTX_REPORT.INDEX_SIZE will also output information on dr$indexname$S table.

Related Topics

"CREATE INDEX" on page 1-36
Table 2–42, " BASIC_STORAGE Attributes" on page 2-76
INDEX_STATS

Creates a report showing various calculated statistics about the text index. This procedure fully scans the text index tables, so it may take a long time to run for large indexes.

Syntax

procedure ctx_report.index_stats(
    index_name     IN VARCHAR2,
    report         IN OUT NOCOPY CLOB,
    part_name      IN VARCHAR2 DEFAULT NULL,
    frag_stats     IN BOOLEAN DEFAULT TRUE,
    list_size      IN NUMBER DEFAULT 100,
    report_format  IN VARCHAR2 DEFAULT FMT_TEXT,
    stat_type      IN VARCHAR2 DEFAULT NULL
);  

index_name
Specify the name of the index to describe. This must be a CONTEXT index.

report
Specify the CLOB locator to which to write the report. If report is NULL, a session-duration temporary CLOB will be created and returned. It is the caller's responsibility to free this temporary CLOB as needed.

The report CLOB will be truncated before report is generated, so any existing contents will be overwritten by this call.

part_name
Specify the name of the index partition. If the index is a local partitioned index, then part_name must be provided. INDEX_STATS will calculate the statistics for that index partition.

frag_stats
Specify TRUE to calculate fragmentation statistics. If frag_stats is FALSE, the report will not show any statistics relating to size of index data. However, the operation should take less time and resources to calculate the token statistics.

list_size
Specify the number of elements in each compiled list. list_size has a maximum value of 1000.

report_format
Specify whether the report should be generated as 'TEXT' or as 'XML'. TEXT is the default. You can also specify the values CTX_REPORT.FMT_TEXT or CTX_REPORT.FMT_XML.

stat_type
Specify the estimated statistics to output. If this parameter is set, then frag_stats is ignored. The possible values are:

<table>
<thead>
<tr>
<th>stat_type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EST_FRAG_STATS</td>
<td>Get the estimated fragmentation stats for the index. When this type is given list_size is ignored.</td>
</tr>
</tbody>
</table>
Example

create table output (result CLOB);

declare
  x clob := null;
begin
  ctx_report.index_stats('tdrbprx21',x);
  insert into output values (x);
  commit;
  dbms_lob.freetemporary(x);
end;
/

set long 32000
set head off
set pagesize 10000
select * from output;

The following sample output is for INDEX_STATS on a context index. This report has been truncated for clarity. It shows some of the token statistics and all of the fragmentation statistics.

The fragmentation statistics are at the end of the report. It tells you optimal row fragmentation, an estimated amount of garbage data in the index, and a list of the most fragmented tokens. Running CTX_DDL.OPTIMIZE_INDEX cleans up the index.

=================================================================
STATISTICS FOR "DR_TEST"."TDRBPRX21"
=================================================================
indexed documents:                                          53
allocated docids:                                          68
$1 rows:                                                16,259

-----------------------------------------------------------------
TOKEN STATISTICS
-----------------------------------------------------------------
unique tokens:                                          13,445
average $1 rows for each token:                           1.21
tokens with most $1 rows:
  telecommunications industry (THEME)                        6
  science and technology (THEME)                             6
  EMAIL (FIELD SECTION "SOURCE")                             6

Note: The estimated statistics for stat_type is only available if query_stats is enabled. See CTX_OUTPUT.ENABLE_QUERY_STATS and CTX_OUTPUT.DISABLE_QUERY_STATS.
INDEX_STATS

DEC (FIELD SECTION "TIMESTAMP") 6
electronic mail (THEME) 6
computer networking (THEME) 6
communications (THEME) 6
95 (FIELD SECTION "TIMESTAMP") 6
15 (FIELD SECTION "TIMESTAMP") 6
HEADLINE (ZONE SECTION) 6

average size for each token: 8
tokens with largest size:
T (NORMAL) 405
SAID (NORMAL) 313
HEADLINE (ZONE SECTION) 272
NEW (NORMAL) 267
I (NORMAL) 230
MILLION (PREFIX) 222
D (NORMAL) 219
MILLION (NORMAL) 215
U (NORMAL) 192
DEC (FIELD SECTION "TIMESTAMP") 186

average frequency for each token: 2.00
most frequent tokens:
HEADLINE (ZONE SECTION) 68
DEC (FIELD SECTION "TIMESTAMP") 62
95 (FIELD SECTION "TIMESTAMP") 62
15 (FIELD SECTION "TIMESTAMP") 62
T (NORMAL) 61
D (NORMAL) 59
881115 (THEME) 58
881115 (NORMAL) 58
I (NORMAL) 55
geography (THEME) 52

token statistics by type:
token type: NORMAL
unique tokens: 6,344
total rows: 7,631
average rows: 1.20
total size: 67,445 (65.86 KB)
average size: 11
average frequency: 2.33
most frequent tokens:
T 61
D 59
881115 58
I 55
SAID 45
C 43
NEW 36
MILLION 32
FIRST 28
COMPANY 27

token type: THEME
unique tokens: 4,563
total rows: 5,523
average rows: 1.21
total size: 21,930 (21.42 KB)
average size: 5

12-10  Oracle Text Reference
average frequency: 2.40
most frequent tokens:
881115 58
political geography 52
geography 52
United States 51
business and economics 50
abstract ideas and concepts 48
North America 48
science and technology 46
NKS 34
nulls 34

The fragmentation portion of this report is as follows:

-----------------------------------------------------------------
FRAGMENTATION STATISTICS
-----------------------------------------------------------------
total size of $I data: 116,772 (114.04 KB)
$I rows: 16,259
estimated $I rows if optimal: 13,445
estimated row fragmentation: 17 %
garbage docids: 15
estimated garbage size: 21,379 (20.88 KB)

most fragmented tokens:
telecommunications industry (THEME) 83 %
science and technology (THEME) 83 %
EMAIL (FIELD SECTION "SOURCE") 83 %
DEC (FIELD SECTION "TIMESTAMP") 83 %
electronic mail (THEME) 83 %
computer networking (THEME) 83 %
communications (THEME) 83 %
95 (FIELD SECTION "TIMESTAMP") 83 %
HEADLINE (ZONE SECTION) 83 %
15 (FIELD SECTION "TIMESTAMP") 83 %

Notes

These metadata are available only when QUERY_STATS is turned on for the index: estimated fragmentation stats, estimated frequently queried tokens, estimated most fragmented frequently queried token, and estimated slowest running queries for the specified index.

CTX_REPORT.INDEX_STATS will also output information on dr$indexname$S table, which is the section data, or SDATA, table.

Related Topics

CTX_OUTPUT.ENABLE_QUERY_STATS on page 10-6
CTX_OUTPUT.DISABLE_QUERY_STATS on page 10-5.
CREATE INDEX on page 1-36.
Table 2–42, "BASIC_STORAGE Attributes" on page 2-76.
Obtain a report of logged queries.

QUERY_LOG_SUMMARY enables you to analyze queries you have logged. For example, suppose you have an application that searches a database of large animals, and your analysis of queries against it shows that users are continually searching for the word *mouse*; this analysis might induce you to rewrite your application so that a search for *mouse* redirects the user to a database for small animals instead of simply returning an unsuccessful search.

With query analysis, you can find out the following:

- Which queries were made
- Which queries were successful
- Which queries were unsuccessful
- How many times each query was made

You can combine these factors in various ways, such as determining the 50 most frequent unsuccessful queries made by your application.

Query logging is begun with `CTX_OUTPUT.START_QUERY_LOG` and terminated with `CTX_OUTPUT.END_QUERY_LOG`.

---

**Note:** You must connect as CTXSYS to use `CTX_REPORT.QUERY_LOG_SUMMARY`.

---

**See Also:** "START_QUERY_LOG" on page 10-16 and "END_QUERY_LOG" on page 10-8

### Syntax

```sql
procedure CTX_REPORT.QUERY_LOG_SUMMARY(
    logfile       IN VARCHAR2,
    indexname     IN VARCHAR2 DEFAULT NULL,
    result_table  IN OUT NOCOPY QUERY_TABLE,
    row_num       IN NUMBER,
    most_freq     IN BOOLEAN DEFAULT TRUE,
    has_hit       IN BOOLEAN DEFAULT TRUE
);```

**logfile**
Specify the name of the logfile that contains the queries.

**indexname**
Specify the name of the context index for which you want the summary report. If you specify `NULL`, the procedure provides a summary report for all context indexes.

**result_table**
Specify the name of the in-memory table of type `TABLE OF RECORD` where the results of the `QUERY_LOG_SUMMARY` are to go. The default is the location specified by the system parameter `LOG_DIRECTORY`.
row_num
The number of rows of results from QUERY_LOG_SUMMARY to be reported into the table named by restab. For example, if this is number is 10, most_freq is TRUE, and has_hit is TRUE, then the procedure returns the 10 most frequent queries that were successful (that is, returned hits).

most_freq
Specify whether QUERY_LOG_SUMMARY should return the most frequent or least frequent queries. The default is most frequent queries. If most_freq is set to FALSE, the procedure returns the least successful queries.

has_hit
Specify whether QUERY_LOG_SUMMARY should return queries that are successful (that is, that generate hits) or unsuccessful queries. The default is to count successful queries; set has_hit to FALSE to return unsuccessful queries.

Example
The following example shows how a query log can be used.

First connect as CTXSYS. Then create and populate two tables, and then create an index for each:

```sql
create table qlogtab1 (tk number primary key, text varchar2(2000));
insert into qlogtab1 values(1, 'The Roman name for France was Gaul.');
insert into qlogtab1 values(2, 'The Tour de France is held each summer.');
insert into qlogtab1 values(3, 'Jacques Anatole Thibault took the pen name Anatole France.');
create index idx_qlog1 on qlogtab1(text) indextype is ctxsys.context;
create table qlogtab2 (tk number primary key, text varchar2(2000));
insert into qlogtab2 values(1, 'The Great Wall of China is about 2400 kilometers long');
insert into qlogtab2 values(2, 'Soccer dates back at least to 217 C.E.');
insert into qlogtab2 values(3, 'The Corn Palace is a tourist attraction in South Dakota.');
create index idx_qlog2 on qlogtab2(text) indextype is ctxsys.context;
```

Turn on query logging, creating a log called query_log:

```sql
exec ctx_output.start_query_log('query.log');
```

Now make some queries (some of which will be unsuccessful):

```sql
select text from qlogtab1 where contains(text, 'France',1)>0;
select text from qlogtab1 where contains(text, 'cheese',1)>0;
select text from qlogtab1 where contains(text, 'Text Wizard',1)>0;
select text from qlogtab2 where contains(text, 'Corn Palace',1)>0;
select text from qlogtab2 where contains(text, 'China',1)>0;
select text from qlogtab2 where contains(text, 'Text Wizards',1)>0;
select text from qlogtab1 where contains(text, 'South Dakota',1)>0;
select text from qlogtab1 where contains(text, 'Text Wizard',1)>0;
select text from qlogtab2 where contains(text, 'company',1)>0;
select text from qlogtab1 where contains(text, 'Text Wizard',1)>0;
select text from qlogtab1 where contains(text, 'database',1)>0;
select text from qlogtab2 where contains(text, 'high-tech',1)>0;
select text from qlogtab1 where contains(text, 'France',1)>0;
select text from qlogtab1 where contains(text, 'database',1)>0;
select text from qlogtab1 where contains(text, 'Japan',1)>0;
select text from qlogtab1 where contains(text, 'Egypt',1)>0;
select text from qlogtab1 where contains(text, 'Argentina',1)>0;
```
select text from qlogtab1 where contains(text, 'Argentina',1)>0;
select text from qlogtab1 where contains(text, 'Argentina',1)>0;
select text from qlogtab1 where contains(text, 'Japan',1)>0;
select text from qlogtab1 where contains(text, 'Egypt',1)>0;
select text from qlogtab1 where contains(text, 'Air Shuttle',1)>0;
select text from qlogtab1 where contains(text, 'Argentina',1)>0;

With the querying over, turn query logging off:

exec ctx_output.end_query_log;

Use QUERY_LOG_SUMMARY to get query reports. In the first instance, you ask to see the three most frequent queries that return successfully. First declare the results table (the_queries).

set serveroutput on;
declare
    the_queries ctx_report.query_table;
begin
    ctx_report.query_log_summary('query.log', null, the_queries,
        row_num=>3, most_freq=>TRUE, has_hit=>TRUE);
dbms_output.put_line('The 3 most frequent queries returning hits');
dbms_output.put_line('number of times  query string');
for i in 1..the_queries.count loop
    dbms_output.put_line(the_queries(i).times||'                '||the_queries(i).query);
end loop;
end;
/

This returns the following:

The 3 most frequent queries returning hits
number of times  query string
3                France
2                China
1                Corn Palace

Next, look for the three most frequent queries on idx_qlog1 that were successful.

declare
    the_queries ctx_report.query_table;
begin
    ctx_report.query_log_summary('query.log', 'idx_qlog1', the_queries,
        row_num=>3, most_freq=>TRUE, has_hit=>TRUE);
dbms_output.put_line('The 3 most frequent queries returning hits for index idx_qlog1');
dbms_output.put_line('number of times  query string');
for i in 1..the_queries.count loop
    dbms_output.put_line(the_queries(i).times||'                '||the_queries(i).query);
end loop;
end;
/

Because only the queries for France were successful, ctx_report.query_log_summary returns the following:

The 3 most frequent queries returning hits for index idx_qlog1
number of times  query string
3                France

Lastly, ask to see the three least frequent queries that returned no hits (that is, queries that were unsuccessful and called infrequently). In this case, you are interested in queries on both context indexes, so you set the indexname parameter to NULL.
declare
    the_queries ctx_report.query_table;
begin
    ctx_report.query_log_summary('query.log', null, the_queries, row_num=>3,
        most_freq=>FALSE, has_hit=>FALSE);
    dbms_output.put_line('The 3 least frequent queries returning no hit');
    dbms_output.put_line('number of times  query string');
    for i in 1..the_queries.count loop
        dbms_output.put_line(the_queries(i).times||'                '||the_queries(i).query);
    end loop;
end;
/

This returns the following results:

The 3 least frequent queries returning no hit
number of times  query string
1                high-tech
1                company
1                cheese

Argentina and Japan do not make this list, because they are queried more than once, while Corn Palace does not make this list because it is successfully queried.
TOKEN_INFO

Creates a report showing the information for a token, decoded. This procedure will fully scan the info for a token, so it may take a long time to run for really large tokens.

You can call this operation as a procedure with an `IN OUT CLOB` parameter or as a function that returns the report as a `CLOB`.

Syntax

```sql
procedure CTX_REPORT.TOKEN_INFO(
    index_name      IN VARCHAR2,
    report          IN OUT NOCOPY CLOB,
    token           IN VARCHAR2,
    token_type      IN NUMBER,
    part_name       IN VARCHAR2 DEFAULT NULL,
    raw_info        IN BOOLEAN  DEFAULT FALSE,
    decoded_info    IN BOOLEAN  DEFAULT TRUE,
    report_format   IN VARCHAR2 DEFAULT FMT_TEXT
);

function CTX_REPORT.TOKEN_INFO(
    index_name      IN VARCHAR2,
    token           IN VARCHAR2,
    token_type      IN NUMBER,
    part_name       IN VARCHAR2 DEFAULT NULL,
    raw_info        IN VARCHAR2 DEFAULT 'N',
    decoded_info    IN VARCHAR2 DEFAULT 'Y',
    report_format   IN VARCHAR2 DEFAULT FMT_TEXT
) return clob;
```

**index_name**
Specify the name of the index.

**report**
Specify the `CLOB` locator to which to write the report.

If report is `NULL`, a session-duration temporary `CLOB` will be created and returned. It is the caller's responsibility to free this temporary `CLOB` as needed.

The `report` `CLOB` will be truncated before report is generated, so any existing contents will be overwritten by this call token may be case-sensitive, depending on the passed-in token type.

**token**
Specify the token text.

**token_type**
Specify the token type. You can use a number returned by the `TOKEN_TYPE` function. `THEME`, `ZONE`, `ATTR`, `PATH`, and `PATH_ATTR` tokens are case-sensitive.

Everything else gets passed through the lexer, so if the index's lexer is case-sensitive, the token input is case-sensitive.

**part_name**
Specify the name of the index partition.

If the index is a local partitioned index, then `part_name` must be provided. `TOKEN_INFO` will apply to just that index partition.
raw_info
Specify TRUE to include a hex dump of the index data. If raw_info is TRUE, the report will include a hex dump of the raw data in the token_info column.

decoded_info
Specify decode and include docid and offset data. If decoded_info is FALSE, CTX_REPORT will not attempt to decode the token information. This is useful when you just want a dump of data.

report_format
Specify whether the report should be generated as 'TEXT' or as 'XML'. TEXT is the default. You can also specify the values CTX_REPORT.FMT_TEXT or CTX_REPORT.FMT_XML.
This is a helper function which translates an English name into a numeric token type. This is suitable for use with \texttt{token\_info}, or any other CTX API which takes in a \texttt{token\_type}.

\begin{verbatim}
function token_type(
    index_name in varchar2,
    type_name  in varchar2
) return number;
\end{verbatim}

\begin{verbatim}
TOKEN_TYPE_TEXT      constant number := 0;
TOKEN_TYPE_THEME     constant number := 1;
TOKEN_TYPE_ZONE_SEC  constant number := 2;
TOKEN_TYPE_ORIG      constant number := 3,
TOKEN_TYPE_ATTR_TEXT constant number := 4;
TOKEN_TYPE_ATTR_SEC  constant number := 5;
TOKEN_TYPE_PREFIX    constant number := 6;
TOKEN_TYPE_PATH_SEC  constant number := 7;
TOKEN_TYPE_PATH_ATTR constant number := 8;
TOKEN_TYPE_STEM      constant number := 9;
\end{verbatim}

\textbf{index\_name}
Specify the name of the index.

\textbf{type\_name}
Specify an English name for \texttt{token\_type}. The following strings are legal input. All input is case-insensitive.

\begin{center}
\begin{tabular}{|l|l|l|}
\hline
\textbf{Input} & \textbf{Meaning} & \textbf{Type Returned} \\
\hline
\texttt{TEXT} & Normal text token. & 0 \\
\texttt{THEME} & Theme token. & 1 \\
\texttt{ZONE SEC} & Zone token. & 2 \\
\texttt{ORIGINAL} & Original form token & 3 \\
\texttt{ATTR TEXT} & Text that occurs in attribute. & 4 \\
\texttt{ATTR SEC} & Attribute section. & 5 \\
\texttt{PREFIX} & Prefix token. & 6 \\
\texttt{PATH SEC} & Path section. & 7 \\
\texttt{PATH ATTR} & Path attribute section. & 8 \\
\texttt{STEM} & Stem form token. & 9 \\
\texttt{FIELD <name> TEXT} & Text token in field section \texttt{<name>} & 16-79 \\
\texttt{FIELD <name> PREFIX} & Prefix token in field section \texttt{<name>} & 616-916 \\
\texttt{FIELD <name> STEM} & Stem token in field section \texttt{<name>} & 916-979 \\
\texttt{NDATA <name>} & NDATA-type token & 200-299 \\
\texttt{TOKEN\_TYPE\_ATTR\_TXT\_PFIX} & Attribute text prefix. & 604 \\
\texttt{TOKEN\_TYPE\_ATTR\_TXT\_STEM} & Attribute text stem. & 904 \\
\hline
\end{tabular}
\end{center}
For \texttt{FIELD} types, the index metadata needs to be read, so if you are going to be calling this a lot for such things, you might want to consider caching the values in local variables rather than calling \texttt{token_type} over and over again.

The constant types (0 - 9) also have constants in this package defined.

Notes

To get token types for \texttt{MDATA} tokens, do not use \texttt{CTX\_REPORT\_TOKEN\_TYPE}; use the \texttt{MDATA} operator instead. (See "\texttt{MDATA}" on page 3-28.) The syntax to use is \texttt{'MDATA fieldname'}.

Example

\begin{verbatim}
typenum := ctx_report.token_type('myindex', 'field author text');
\end{verbatim}
This chapter provides reference information for using the CTX_THES package to manage and browse thesauri. These thesaurus functions are based on the ISO-2788 and ANSI Z39.19 standards except where noted.

Knowing how information is stored in your thesaurus helps in writing queries with thesaurus operators. You can also use a thesaurus to extend the knowledge base, which is used for ABOUT queries in English and French and for generating document themes.

CTX_THES contains the following stored procedures and functions:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALTER_PHRASE</td>
<td>Alters thesaurus phrase.</td>
</tr>
<tr>
<td>ALTER_THESAURUS</td>
<td>Renames or truncates a thesaurus.</td>
</tr>
<tr>
<td>BT</td>
<td>Returns all broader terms of a phrase.</td>
</tr>
<tr>
<td>BTG</td>
<td>Returns all broader terms generic of a phrase.</td>
</tr>
<tr>
<td>BTI</td>
<td>Returns all broader terms instance of a phrase.</td>
</tr>
<tr>
<td>BTP</td>
<td>Returns all broader terms partitive of a phrase.</td>
</tr>
<tr>
<td>CREATE_PHRASE</td>
<td>Adds a phrase to the specified thesaurus.</td>
</tr>
<tr>
<td>CREATE_RELATION</td>
<td>Creates a relation between two phrases.</td>
</tr>
<tr>
<td>CREATE_THESAURUS</td>
<td>Creates the specified thesaurus.</td>
</tr>
<tr>
<td>CREATE_TRANSLATION</td>
<td>Creates a new translation for a phrase.</td>
</tr>
<tr>
<td>DROP_PHRASE</td>
<td>Removes a phrase from thesaurus.</td>
</tr>
<tr>
<td>DROP_RELATION</td>
<td>Removes a relation between two phrases.</td>
</tr>
<tr>
<td>DROP_THESAURUS</td>
<td>Drops the specified thesaurus from the thesaurus tables.</td>
</tr>
<tr>
<td>DROP_TRANSLATION</td>
<td>Drops a translation for a phrase.</td>
</tr>
<tr>
<td>HAS_RELATION</td>
<td>Tests for the existence of a thesaurus relation.</td>
</tr>
<tr>
<td>NT</td>
<td>Returns all narrower terms of a phrase.</td>
</tr>
<tr>
<td>NTG</td>
<td>Returns all narrower terms generic of a phrase.</td>
</tr>
<tr>
<td>NTI</td>
<td>Returns all narrower terms instance of a phrase.</td>
</tr>
<tr>
<td>NTP</td>
<td>Returns all narrower terms partitive of a phrase.</td>
</tr>
<tr>
<td>Name</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>RT</td>
<td>Returns the related terms of a phrase</td>
</tr>
<tr>
<td>SN</td>
<td>Returns scope note for phrase.</td>
</tr>
<tr>
<td>SYN</td>
<td>Returns the synonym terms of a phrase</td>
</tr>
<tr>
<td>THES_TT</td>
<td>Returns all top terms for phrase.</td>
</tr>
<tr>
<td>TR</td>
<td>Returns the foreign equivalent of a phrase.</td>
</tr>
<tr>
<td>TRSYN</td>
<td>Returns the foreign equivalent of a phrase, synonyms of the phrase, and foreign equivalent of the synonyms.</td>
</tr>
<tr>
<td>TT</td>
<td>Returns the top term of a phrase.</td>
</tr>
<tr>
<td>UPDATE_TRANSLATION</td>
<td>Updates an existing translation.</td>
</tr>
</tbody>
</table>

**See Also:** Chapter 3, "Oracle Text CONTAINS Query Operators" for more information about the thesaurus operators.
ALTER_PHRASE

Alters an existing phrase in the thesaurus. Only CTXSYS or thesaurus owner can alter a phrase.

Syntax

```sql
CTX_THES.ALTER_PHRASE(tname in varchar2,
                         phrase in varchar2,
                         op in varchar2,
                         operand in varchar2 default null);
```

tname
Specify the thesaurus name.

phrase
Specify a phrase to alter.

op
Specify the alter operation as a string or symbol. You can specify one of the following operations with the op and operand pair:

<table>
<thead>
<tr>
<th>op</th>
<th>meaning</th>
<th>operand</th>
</tr>
</thead>
<tbody>
<tr>
<td>RENAME or CTX_THES.OP_RENAME</td>
<td>Rename phrase. If the new phrase already exists in the thesaurus, this procedure raises an exception.</td>
<td>Specify a new phrase. You can include qualifiers to change, add, or remove qualifiers from phrases.</td>
</tr>
<tr>
<td>PT or CTX_THES.OP_PT</td>
<td>Make phrase the preferred term. Existing preferred terms in the synonym ring becomes non-preferred synonym.</td>
<td>(none)</td>
</tr>
<tr>
<td>SN or CTX_THES.OP_SN</td>
<td>Change the scope note on the phrase.</td>
<td>Specify a new scope note.</td>
</tr>
</tbody>
</table>

Examples

Correct misspelled word in thesaurus:
```sql
ctx_thes.alter_phrase('thes1', 'tee', 'rename', 'tea');
```

Remove qualifier from mercury (metal):
```sql
ctx_thes.alter_phrase('thes1', 'mercury (metal)', 'rename', 'mercury');
```

Add qualifier to mercury:
```sql
ctx_thes.alter_phrase('thes1', 'mercury', 'rename', 'mercury (planet)');
```

Make Kowalski the preferred term in its synonym ring:
```sql
ctx_thes.alter_phrase('thes1', 'Kowalski', 'pt');
```
Change scope note for view cameras:

ctx_thes.alter_phrase('thes1', 'view cameras', 'sn', 'Cameras with lens focusing');
ALTER_THESAURUS

Use this procedure to rename or truncate an existing thesaurus. Only the thesaurus owner or CTXSYS can invoke this function on a given thesaurus.

Syntax

```
CTX_THES.ALTER_THESAURUS(tname     in   varchar2,
                           op       in   varchar2,
                           operand  in   varchar2 default null);
```

tname
Specify the thesaurus name.

op
Specify the alter operation as a string or symbol. You can specify one of two operations:

<table>
<thead>
<tr>
<th>op</th>
<th>Meaning</th>
<th>operand</th>
</tr>
</thead>
<tbody>
<tr>
<td>RENAME</td>
<td>Rename thesaurus. Returns an error if the new name already exists.</td>
<td>Specify a new thesaurus name.</td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTX_THES.OP_RENAME</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRUNCATE</td>
<td>Truncate thesaurus.</td>
<td>None.</td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTX_THES.OP_TRUNCATE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

operand
Specify the argument to the alter operation. See table for op.

Examples

Rename thesaurus THES1 to MEDICAL:

```
ctx_thes.alter_thesaurus('thes1', 'rename', 'medical');
```

or

```
ctx_thes.alter_thesaurus('thes1', ctx_thes.op_rename, 'medical');
```

You can use symbols for any op argument, but all further examples will use strings.

Remove all phrases and relations from thesaurus THES1:

```
ctx_thes.alter_thesaurus('thes1', 'trunc') ;
```
This function returns all broader terms of a phrase as recorded in the specified thesaurus.

**Syntax 1: Table Result**

```
CTX_THES.BT(restab IN OUT NOCOPY EXP_TAB,
    phrase IN VARCHAR2,
    lvl IN NUMBER DEFAULT 1,
    tname IN VARCHAR2 DEFAULT 'DEFAULT');
```

**Syntax 2: String Result**

```
CTX_THES.BT(phrase IN VARCHAR2,
    lvl IN NUMBER DEFAULT 1,
    tname IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN VARCHAR2;
```

**restab**
Optionally, specify the name of the expansion table to store the results. This table must be of type `EXP_TAB` which the system defines as follows:

```
type exp_rec is record {
    xrel varchar2(12),
    xlevel number,
    xphrase varchar2(256)
};
type exp_tab is table of exp_rec index by binary_integer;
```

**See Also:** "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about `EXP_TAB`

**phrase**
Specify a phrase to lookup in thesaurus.

**lvl**
Specify how many levels of broader terms to return. For example 2 means get the broader terms of the broader terms of the phrase.

**tname**
Specify a thesaurus name. If not specified, system default thesaurus is used.

**Returns**
This function returns a string of broader terms in the form:

```
{bt1}|{bt2}|{bt3} ...
```

**Example**

**String Result**
Consider a thesaurus named `MY_THES` that has an entry for `cat` as follows:

```
cat
    BT1 feline
```
To look up the broader terms for *cat* up to two levels, enter the following statements:

```sql
set serveroutput on
declare
terms varchar2(2000);
begin
terms := ctx_thes.bt('CAT', 2, 'MY_THES');
dbms_output.put_line('The broader expansion for CAT is: '||terms);
end;
```

This code produces the following output:

The broader expansion for CAT is: {cat}|{feline}|{mammal}

### Table Result

The following code does an broader term lookup for *white wolf* using the table result:

```sql
set serveroutput on
declare
  xtab ctx_thes.exp_tab;
begin
ctx_thes.bt(xtab, 'white wolf', 2, 'my_thesaurus');
for i in 1..xtab.count loop
dbms_output.put_line(xtab(i).rel||' '||xtab(i).phrase);
end loop;
end;
```

This code produces the following output:

<table>
<thead>
<tr>
<th>PHRASE</th>
<th>WHITE WOLF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT</td>
<td>WOLF</td>
</tr>
<tr>
<td>BT</td>
<td>CANINE</td>
</tr>
<tr>
<td>BT</td>
<td>ANIMAL</td>
</tr>
</tbody>
</table>

### Related Topics

**OUTPUT_STYLE**

Broader Term (BT, BTG, BTP, BTI) Operators in Chapter 3, *Oracle Text CONTAINS Query Operators*
This function returns all broader terms generic of a phrase as recorded in the specified thesaurus.

Syntax 1: Table Result

```
CTX_THES.BTG(restab IN OUT NOCOPY EXP_TAB,
             phrase IN VARCHAR2,
             lvl IN NUMBER DEFAULT 1,
             tname IN VARCHAR2 DEFAULT 'DEFAULT');
```

Syntax 2: String Result

```
CTX_THES.BTG(phrase IN VARCHAR2,
             lvl IN NUMBER DEFAULT 1,
             tname IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN VARCHAR2;
```

`restab`
Optionally, specify the name of the expansion table to store the results. This table must be of type `EXP_TAB` which the system defines as follows:

```
type exp_rec is record {
  xrel varchar2(12),
  xlevel number,
  xphrase varchar2(256)
};
type exp_tab is table of exp_rec index by binary_integer;
```

See Also: "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about `EXP_TAB`.

`phrase`
Specify a phrase to lookup in thesaurus.

`lvl`
Specify how many levels of broader terms to return. For example 2 means get the broader terms of the broader terms of the phrase.

`tname`
Specify thesaurus name. If not specified, system default thesaurus is used.

Returns

This function returns a string of broader terms generic in the form:

```
{bt1}|{bt2}|{bt3} ...
```

Example

To look up the broader terms generic for `cat` up to two levels, enter the following statements:

```
set serveroutput on
declare
  terms varchar2(2000);
begin
```
terms := ctx_thes.btg('CAT', 2, 'MY_THES');
dbms_output.put_line('the broader expansion for CAT is: ' || terms);
end;

Related Topics

OUTPUT_STYLE

Broader Term (BT, BTG, BTP, BTI) Operators in Chapter 3, "Oracle Text CONTAINS Query Operators"
This function returns all broader terms instance of a phrase as recorded in the specified thesaurus.

Syntax 1: Table Result

```
CTX_THES.BTI(restab IN OUT NOCOPY EXP_TAB,
             phrase IN VARCHAR2,
             lvl    IN NUMBER DEFAULT 1,
             tname  IN VARCHAR2 DEFAULT 'DEFAULT');
```

Syntax 2: String Result

```
CTX_THES.BTI(phrase IN VARCHAR2,
             lvl    IN NUMBER DEFAULT 1,
             tname  IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN VARCHAR2;
```

restab
Optionally, specify the name of the expansion table to store the results. This table must be of type EXP_TAB which the system defines as follows:

```
type exp_rec is record {
  xrel varchar2(12),
  xlevel number,
  xphrase varchar2(256)
};
type exp_tab is table of exp_rec index by binary_integer;
```

See Also: "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about EXP_TAB.

phrase
Specify a phrase to lookup in thesaurus.

lvl
Specify how many levels of broader terms to return. For example 2 means get the broader terms of the broader terms of the phrase.

tname
Specify a thesaurus name. If not specified, system default thesaurus is used.

Returns
This function returns a string of broader terms instance in the form:

```
{bt1}|{bt2}|{bt3} ...
```

Example
To look up the broader terms instance for cat up to two levels, enter the following statements:

```
set serveroutput on
declare
  terms varchar2(2000);
begin
```
terms := ctx_thes.bti('CAT', 2, 'MY_THES');
dbms_output.put_line('the broader expansion for CAT is: '||terms);
end;

Related Topics

OUTPUT_STYLE

Broader Term (BT, BTG, BTP, BTI) Operators in Chapter 3, "Oracle Text CONTAINS Query Operators"
This function returns all broader terms partitive of a phrase as recorded in the specified thesaurus.

**Syntax 1: Table Result**

```sql
CTX_THES.BTP(restab IN OUT NOCOPY EXP_TAB,
             phrase IN VARCHAR2,
             lvl IN NUMBER DEFAULT 1,
             tname IN VARCHAR2 DEFAULT 'DEFAULT');
```

**Syntax 2: String Result**

```sql
CTX_THES.BTP(phrase IN VARCHAR2,
             lvl IN NUMBER DEFAULT 1,
             tname IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN VARCHAR2;
```

**restab**
Optionally, specify the name of the expansion table to store the results. This table must be of type `EXP_TAB` which the system defines as follows:

```sql
type exp_rec is record (
  xrel varchar2(12),
  xlevel number,
  xphrase varchar2(256)
);
type exp_tab is table of exp_rec index by binary_integer;
```

**See Also:** "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about `EXP_TAB`.

**phrase**
Specify a phrase to lookup in thesaurus.

**lvl**
Specify how many levels of broader terms to return. For example 2 means get the broader terms of the broader terms of the phrase.

**tname**
Specify a thesaurus name. If not specified, the system default thesaurus is used.

**Returns**
This function returns a string of broader terms in the form:

```
{bt1}|{bt2}|{bt3} ...
```

**Example**
To look up the two broader terms partitive for `cat`, enter the following statements:

```sql
declare
terms varchar2(2000);
begin
terms := ctx_thes.btp('CAT', 2, 'MY_THES');
dbms_output.put_line('the broader expansion for CAT is: '||terms);
```
Related Topics

OUTPUT_STYLE
Broader Term (BT, BTG, BTP, BTI) Operators in Chapter 3, "Oracle Text CONTAINS Query Operators"
CREATE_PHRASE

The CREATE_PHRASE procedure adds a new phrase to the specified thesaurus.

Note: Even though you can create thesaurus relations with this procedure, Oracle recommends that you use CTX_THES.CREATE_RELATION rather than CTX_THES.CREATE_PHRASE to create relations in a thesaurus.

Syntax

```
CTX_THES.CREATE_PHRASE(tname   IN VARCHAR2,
                          phrase  IN VARCHAR2,
                          rel     IN VARCHAR2 DEFAULT NULL,
                          relname IN VARCHAR2 DEFAULT NULL);
```

tname
Specify the name of the thesaurus in which the new phrase is added or the existing phrase is located.

phrase
Specify the phrase to be added to a thesaurus or the phrase for which a new relationship is created.

rel
Specify the new relationship between phrase and relname. This parameter is supported only for backward compatibility. Use CTX_THES.CREATE_RELATION to create new relations in a thesaurus.

relname
Specify the existing phrase that is related to phrase. This parameter is supported only for backward compatibility. Use CTX_THES.CREATE_RELATION to create new relations in a thesaurus.

Returns

The ID for the entry.

Example

In this example, two new phrases (os and operating system) are created in a thesaurus named tech_thes.

```
begin
  ctx_thes.create_phrase('tech_thes','os');
  ctx_thes.create_phrase('tech_thes','operating system');
end;
```
CREATE_RELATION

Creates a relation between two phrases in the thesaurus. The synonym ring is limited in length to about 4000 synonyms, depending on word length.

---

**Note:** Oracle recommends that you use `CTX_THES.CREATE_RELATION` rather than `CTX_THES.CREATE_PHRASE` to create relations in a thesaurus.

---

Only thesaurus owner and `CTXSYS` can invoke this procedure on a given thesaurus.

**Syntax**

```sql
CTX_THES.CREATE_RELATION(tname in varchar2, phrase in varchar2, rel in varchar2, relphrase in varchar2);
```

**tname**
Specify the thesaurus name

**phrase**
Specify the phrase to alter or create. If `phrase` is a disambiguated homograph, you must specify the qualifier. If `phrase` does not exist in the thesaurus, it is created.

**rel**
Specify the relation to create. The relation is from `phrase` to `relphrase`. You can specify one of the following relations:

<table>
<thead>
<tr>
<th>relation</th>
<th>meaning</th>
<th>relphrase</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>BT*/NT*</code></td>
<td>Add hierarchical relation.</td>
<td>Specify the related phrase. The relationship is interpreted from phrase to relphrase.</td>
</tr>
<tr>
<td><code>RT</code></td>
<td>Add associative relation.</td>
<td>Specify the phrase to associate.</td>
</tr>
<tr>
<td><code>SYN</code></td>
<td>Add phrase to a synonym ring.</td>
<td>Specify an existing phrase in the synonym ring.</td>
</tr>
</tbody>
</table>

**relphrase**
Specify the related phrase. If `relphrase` does not exist in `tname`, `relphrase` is created. See table for `rel`.

**Notes**

The relation you specify for `rel` is interpreted as from phrase to relphrase. For example, consider dog with broader term animal:

- `dog`
  - `BT animal`

To add this relation, specify the arguments as follows:

```sql
begin
```

...
CREATE_RELATION

CTX_THES.CREATe_RELATION('thes', 'dog', 'BT', 'animal');
end;

Note: The order in which you specify arguments for CTX_THES.CREATe_RELATION is different from the order you specify them with CTX_THES.CREATe_PHRASE.

Examples

Create relation VEHICLE NT CAR:
ctx_thes.create_relation('thes1', 'vehicle', 'NT', 'car');

Create Japanese translation for you:
ctx_thes.create_relation('thes1', 'you', 'JAPANESE:', 'kimi');
The `CREATE_THESAURUS` procedure creates an empty thesaurus with the specified name in the thesaurus tables.

**Syntax**

```sql
CTX_THES.CREATE_THESAURUS (name           IN VARCHAR2,
                           casesens       IN BOOLEAN DEFAULT FALSE);
```

**name**

Specify the name of the thesaurus to be created. The name of the thesaurus must be unique. If a thesaurus with the specified name already exists, `CREATE_THESAURUS` returns an error and does not create the thesaurus.

**casesens**

Specify whether the thesaurus to be created is case-sensitive. If `casesens` is `true`, Oracle Text retains the cases of all terms entered in the specified thesaurus. As a result, queries that use the thesaurus are case-sensitive.

**Example**

```sql
begin
  ctx_thes.create_thesaurus ('tech_thes', FALSE);
end;
```
CREATE_TRANSLATION

Use this procedure to create a new translation for a phrase in a specified language.

Syntax

```sql
CTX_THES.CREATE_TRANSLATION(tname in varchar2,
                                 phrase in varchar2,
                                 language in varchar2,
                                 translation in varchar2);
```

**tname**
Specify the name of the thesaurus, using no more than 30 characters.

**phrase**
Specify the phrase in the thesaurus to which to add a translation. Phrase must already exist in the thesaurus, or an error is raised.

**language**
Specify the language of the translation, using no more than 10 characters.

**translation**
Specify the translated term, using no more than 256 characters.

If a translation for this phrase already exists, this new translation is added without removing that original translation, so long as that original translation is not the same. Adding the same translation twice results in an error.

**Example**

The following code adds the Spanish translation for `dog` to `my_thes`:

```sql
begin
    ctx_thes.create_translation('my_thes', 'dog', 'SPANISH', 'PERRO');
end;
```
DROP_PHRASE

Removes a phrase from the thesaurus. Only thesaurus owner and CTXSYS can invoke this procedure on a given thesaurus.

Syntax

```sql
CTX_THESES.DROP_PHRASE(tname in varchar2,
                          phrase in varchar2);
```

**tname**
Specify thesaurus name.

**phrase**
Specify a phrase to drop. If the phrase is a disambiguated homograph, then you must include the qualifier. If the phrase does not exist in tname, then this procedure raises an exception.

BT*/ NT* relations are patched around the dropped phrase. For example, if A has a BT B, and B has BT C, after B is dropped, A has BT C.

When a word has multiple broader terms, then a relationship is established for each narrower term to each broader term.

Note that BT, BTG, BTP, and BTI are separate hierarchies, so if A has BTG B, and B has BTI C, when B is dropped, there is no relation implicitly created between A and C.

RT relations are not patched. For example, if A has RT B, and B has RT C, then if B is dropped, there is no associative relation created between A and C.

Example

Assume you have the following relations defined in *mythes*:

```plaintext
wolf
   BT canine
canine
   BT animal
```

You drop phrase *canine*:

```sql
begin
    ctx_thes.drop_phrase('mythes', 'canine');
end;
```

The resulting thesaurus is patched and looks like:

```plaintext
wolf
   BT animal
```
DROP_RELATION

Removes a relation between two phrases from the thesaurus.

**Note:** `CTX_THES.DROP_RELATION` removes only the relation between two phrases. Phrases are never removed by this call.

Only thesaurus owner and CTXSYS can invoke this procedure on a given thesaurus.

**Syntax**

```sql
CTX_THES.DROP_RELATION(tname in varchar2,
                        phrase in varchar2,
                        rel in varchar2,
                        relphrase in varchar2 default null);
```

**tname**
Specify the thesaurus name.

**phrase**
Specify the filing phrase.

**rel**
Specify the relation to drop. The relation is from `phrase` to `relphrase`. You can specify one of the following relations:

<table>
<thead>
<tr>
<th>relation</th>
<th>meaning</th>
<th>relphrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT*/NT*</td>
<td>Remove hierarchical relation.</td>
<td>Optional specify relphrase. If not provided, all relations of that type for the phrase are removed.</td>
</tr>
<tr>
<td>RT</td>
<td>Remove associative relation.</td>
<td>Optionally specify relphrase. If not provided, all RT relations for the phrase are removed.</td>
</tr>
<tr>
<td>SYN</td>
<td>Remove phrase from its synonym ring.</td>
<td>(none)</td>
</tr>
<tr>
<td>PT</td>
<td>Remove preferred term designation from the phrase. The phrase remains in the synonym ring.</td>
<td>(none)</td>
</tr>
<tr>
<td>language</td>
<td>Remove a translation from a phrase.</td>
<td>Optionally specify relphrase. You can specify relphrase when there are multiple translations for a phrase for the language, and you want to remove just one translation. If relphrase is NULL, all translations for the phrase for the language are removed.</td>
</tr>
</tbody>
</table>

**relphrase**
Specify the related phrase.
Notes

The relation you specify for rel is interpreted as from phrase to relphrase. For example, consider dog with broader term animal:

dog
   BT animal

To remove this relation, specify the arguments as follows:

begin
  CTX_THES.DROP_RELATION('thes','dog','BT','animal');
end;

You can also remove this relation using NT as follows:

begin
  CTX_THES.DROP_RELATION('thes','animal','NT','dog');
end;

Example

Remove relation VEHICLE NT CAR:

ctx_thes.drop_relation('thes1', 'vehicle', 'NT', 'car');

Remove all narrower term relations for vehicle:

ctx_thes.drop_relation('thes1', 'vehicle', 'NT');

Remove Japanese translations for me:

ctx_thes.drop_relation('thes1', 'me', 'JAPANESE:');

Remove a specific Japanese translation for me:

ctx_thes.drop_relation('thes1', 'me', 'JAPANESE:', 'boku')
The DROP_THESAURUS procedure deletes the specified thesaurus and all of its entries from the thesaurus tables.

Syntax

`CTX_THES.DROP_THESAURUS(name IN VARCHAR2);`

**name**
Specify the name of the thesaurus to be dropped.

Example

```sql
begin
    ctx_thes.drop_thesaurus('tech_thes');
end;
```
DROP_TRANSLATION

Use this procedure to remove one or more translations for a phrase.

Syntax

```
CTX THES.DROP_TRANSLATION (tname in varchar2,
phrase in varchar2,
language in varchar2 default null,
translation in varchar2 default null);
```

tname
Specify the name of the thesaurus, using no more than 30 characters.

phrase
Specify the phrase in the thesaurus to which to remove a translation. The phrase must already exist in the thesaurus or an error is raised.

language
Optionally, specify the language of the translation, using no more than 10 characters. If not specified, the translation must also not be specified and all translations in all languages for the phrase are removed. An error is raised if the phrase has no translations.

translation
Optionally, specify the translated term to remove, using no more than 256 characters. If no such translation exists, an error is raised.

Example

The following code removes the Spanish translation for dog:

```
begin
  ctx_thes.drop_translation('my_thes', 'dog', 'SPANISH', 'PERRO');
end;
```

To remove all translations for dog in all languages:

```
begin
  ctx_thes.drop_translation('my_thes', 'dog');
end;
```
HAS_RELATION

Use this procedure to test that a thesaurus relation exists without actually doing the expansion. The function returns TRUE if the phrase has any of the relations in the specified list.

Syntax

CTX.THES.HAS_RELATION(phrase in varchar2,
            rel in varchar2,
            tname in varchar2 default 'DEFAULT')
            returns boolean;

   phrase
   Specify the phrase.

   rel
   Specify a single thesaural relation or a comma-delimited list of relations, except PT. Specify 'ANY' for any relation.

   tname
   Specify the thesaurus name.

Example

The following example returns TRUE if the phrase cat in the DEFAULT thesaurus has any broader terms or broader generic terms:

set serveroutput on
result boolean;
begin
   result := ctx_thes.has_relation('cat','BT,BTG');
   if (result) then dbms_output.put_line('TRUE');
   else  dbms_output.put_line('FALSE');
end if;
end;
This function returns all narrower terms of a phrase as recorded in the specified thesaurus.

**Syntax 1: Table Result**

```
CTX_THES.NT(restab IN OUT NOCOPY EXP_TAB,
    phrase IN VARCHAR2,
    lvl    IN NUMBER DEFAULT 1,
    tname  IN VARCHAR2 DEFAULT 'DEFAULT');
```

**Syntax 2: String Result**

```
CTX_THES.NT(phrase IN VARCHAR2,
    lvl    IN NUMBER DEFAULT 1,
    tname  IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN VARCHAR2;
```

**restab**
Optionally, specify the name of the expansion table to store the results. This table must be of type `EXP_TAB` which the system defines as follows:

```sql
type exp_rec is record {
    xrel varchar2(12),
    xlevel number,
    xphrase varchar2(256)
};
type exp_tab is table of exp_rec index by binary_integer;
```

**See Also:** "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about EXP_TAB.

**phrase**
Specify a phrase to lookup in thesaurus.

**lvl**
Specify how many levels of narrower terms to return. For example 2 means get the narrower terms of the narrower terms of the phrase.

**tname**
Specify thesaurus name. If not specified, system default thesaurus is used.

**Returns**
This function returns a string of narrower terms in the form:

```
(nt1)(nt2)(nt3) ...
```

**Example**

**String Result**
Consider a thesaurus named MY_THES that has an entry for `cat` as follows:

```
cat
NT domestic cat
```
To look up the narrower terms for cat down to two levels, enter the following statements:

```sql
declare
    terms varchar2(2000);
begin
    terms := ctx_thes.nt('CAT', 2, 'MY_THES');
    dbms_output.put_line('the narrower expansion for CAT is: '||terms);
end;
```

This code produces the following output:

```
the narrower expansion for CAT is: {cat}|{domestic cat}|{Persian cat}|{Siamese cat}| {wild cat}
```

### Table Result

The following code does an narrower term lookup for canine using the table result:

```sql
declare
    xtab ctx_thes.exp_tab;
begin
    ctx_thes.nt(xtab, 'canine', 2, 'my_thesaurus');
    for i in 1..xtab.count loop
        dbms_output.put_line(lpad(' ', 2*xtab(i).xlevel) ||
            xtab(i).xrel || ' ' || xtab(i).xphrase);
    end loop;
end;
```

This code produces the following output:

```
PHRASE CANINE
NT WOLF (Canis lupus)
  NT WHITE WOLF
  NT GREY WOLF
NT DOG (Canis familiaris)
  NT PIT BULL
  NT DASCHUND
  NT CHIHUAHUA
NT HYENA (Canis mesomelas)
NT COYOTE (Canis latrans)
```

### Related Topics

**OUTPUT_STYLE**

Narrower Term (NT, NTG, NTP, NTI) Operators in Chapter 3, "Oracle Text CONTAINS Query Operators"
This function returns all narrower terms generic of a phrase as recorded in the specified thesaurus.

Syntax 1: Table Result

```
CTX_THES.NTG(restab IN OUT NOCOPY EXP_TAB,
phrase IN VARCHAR2,
lvl    IN NUMBER DEFAULT 1,
tname  IN VARCHAR2 DEFAULT 'DEFAULT');
```

Syntax 2: String Result

```
CTX_THES.NTG(phrase IN VARCHAR2,
lvl    IN NUMBER DEFAULT 1,
tname  IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN VARCHAR2;
```

**restab**

Optionally, specify the name of the expansion table to store the results. This table must be of type EXP_TAB which the system defines as follows:

```
type exp_rec is record {
    xrel varchar2(12),
    xlevel number,
    xphrase varchar2(256)
};
type exp_tab is table of exp_rec index by binary_integer;
```

**See Also:** "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about EXP_TAB.

**phrase**

Specify a phrase to lookup in thesaurus.

**lvl**

Specify how many levels of narrower terms to return. For example 2 means get the narrower terms of the narrower terms of the phrase.

**tname**

Specify the thesaurus name. If not specified, system default thesaurus is used.

**Returns**

This function returns a string of narrower terms generic in the form:

```
(nt1)|(nt2)|(nt3) ...
```

**Example**

To look up the narrower terms generic for *cat* down to two levels, enter the following statements:

```
declare
    terms varchar2(2000);
begin
    terms := ctx_thes.ntg('CAT', 2, 'MY_THES');
```

dbms_output.put_line('the narrower expansion for CAT is: ||terms);
end;

Related Topics

OUTPUT_STYLE
Narrower Term (NT, NTG, NTP, NTI) Operators in Chapter 3, "Oracle Text CONTAINS Query Operators"
NTI

This function returns all narrower terms instance of a phrase as recorded in the specified thesaurus.

Syntax 1: Table Result

```sql
CTX_THES.NTI(restab IN OUT NOCOPY EXP_TAB,
    phrase IN VARCHAR2,
    lvl    IN NUMBER DEFAULT 1,
    tname  IN VARCHAR2 DEFAULT 'DEFAULT');
```

Syntax 2: String Result

```sql
CTX_THES.NTI(phrase IN VARCHAR2,
    lvl    IN NUMBER DEFAULT 1,
    tname  IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN VARCHAR2;
```

restab
Optionally, specify the name of the expansion table to store the results. This table must be of type EXP_TAB which the system defines as follows:

```sql
type exp_rec is record {
    xrel varchar2(12),
    xlevel number,
    xphrase varchar2(256)
};
type exp_tab is table of exp_rec index by binary_integer;
```

See Also: "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about EXP_TAB.

phrase
Specify a phrase to lookup in thesaurus.

lvl
Specify how many levels of narrower terms to return. For example 2 means get the narrower terms of the narrower terms of the phrase.

tname
Specify the thesaurus name. If not specified, system default thesaurus is used.

Returns

This function returns a string of narrower terms instance in the form:

```
(nt1)|(nt2)|(nt3) ... 
```

Example

To look up the narrower terms instance for cat down to two levels, enter the following statements:

```sql
declare
    terms varchar2(2000);
begin
    terms := ctx_thes.nti('CAT', 2, 'MY_THES');
```
dbms_output.put_line('the narrower expansion for CAT is: '||terms);
end;

Related Topics

OUTPUT_STYLE

Narrower Term (NT, NTG, NTP, NTI) Operators in Chapter 3, "Oracle Text CONTAINS Query Operators"
NTP

This function returns all narrower terms partitive of a phrase as recorded in the specified thesaurus.

Syntax 1: Table Result

```sql
CTX_THES.NTP(restab IN OUT NOCOPY EXP_TAB,
    phrase IN VARCHAR2,
    lvl    IN NUMBER DEFAULT 1,
    tname  IN VARCHAR2 DEFAULT 'DEFAULT');
```

Syntax 2: String Result

```sql
CTX_THES.NTP(phrase IN VARCHAR2,
    lvl    IN NUMBER DEFAULT 1,
    tname  IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN VARCHAR2;
```

**restab**

Optionally, specify the name of the expansion table to store the results. This table must be of type `EXP_TAB` which the system defines as follows:

```sql
type exp_rec is record {
    xrel varchar2(12),
    xlevel number,
    xphrase varchar2(256)
};
type exp_tab is table of exp_rec index by binary_integer;
```

**See Also:** "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about `EXP_TAB`.

**phrase**

Specify a phrase to lookup in thesaurus.

**lvl**

Specify how many levels of narrower terms to return. For example 2 means get the narrower terms of the narrower terms of the phrase.

**tname**

Specify the thesaurus name. If not specified, system default thesaurus is used.

**Returns**

This function returns a string of narrower terms partitive in the form:

```
{nt1}|{nt2}|{nt3} ...
```

**Example**

To look up the narrower terms partitive for `cat` down to two levels, enter the following statements:

```sql
declare
    terms varchar2(2000);
begin
    terms := ctx_thes.ntp('CAT', 2, 'MY_THES');
```


```sql
    dbms_output.put_line('the narrower expansion for CAT is: '||terms);
end;
```

**Related Topics**

OUTPUT_STYLE
Narrower Term (NT, NTG, NTP, NTI) Operators in Chapter 3, "Oracle Text CONTAINS Query Operators"
OUTPUT_STYLE

Sets the output style for the return string of the CTX_THES expansion functions. This procedure has no effect on the table results to the CTX_THES expansion functions.

Syntax

```
CTX_THES.OUTPUT_STYLE (showlevel IN BOOLEAN DEFAULT FALSE, showqualify IN BOOLEAN DEFAULT FALSE, showpt IN BOOLEAN DEFAULT FALSE, showid IN BOOLEAN DEFAULT FALSE);
```

- **showlevel**
  Specify TRUE to show level in BT/NT expansions.

- **showqualify**
  Specify TRUE to show phrase qualifiers.

- **showpt**
  Specify TRUE to show preferred terms with an asterisk *.

- **showid**
  Specify TRUE to show phrase ids.

Notes

The general syntax of the return string for CTX_THES expansion functions is:

```
{pt indicator:phrase (qualifier):level:phraseid}
```

Preferred term indicator is an asterisk then a colon at the start of the phrase. The qualifier is in parentheses after a space at the end of the phrase. Level is a number.

The following is an example return string for turkey the bird:

```
*:TURKEY (BIRD):1:1234
```
This function returns the preferred term of a phrase as recorded in the specified thesaurus.

Syntax 1: Table Result

```sql
CTX_THES.PT(restab IN OUT NOCOPY EXP_TAB,
            phrase IN VARCHAR2,
            tname IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN varchar2;
```

Syntax 2: String Result

```sql
CTX_THES.PT(phrase IN VARCHAR2,
            tname IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN varchar2;
```

**restab**
Optionally, specify the name of the expansion table to store the results. This table must be of type `EXP_TAB` which the system defines as follows:

```sql
type exp_rec is record (
    xrel varchar2(12),
    xlevel number,
    xphrase varchar2(256)
);
type exp_tab is table of exp_rec index by binary_integer;
```

**See Also:** "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about `EXP_TAB`.

**phrase**
Specify a phrase to lookup in thesaurus.

**tname**
Specify thesaurus name. If not specified, system default thesaurus is used.

**Returns**
This function returns the preferred term as a string in the form:

```
{pt}
```

**Example**
Consider a thesaurus `MY_THES` with the following preferred term definition for automobile:

```
AUTOMOBILE
    PT CAR
```

To look up the preferred term for `automobile`, execute the following code:

```sql
declare
terms varchar2(2000);
begin
terms := ctx_thes.pt('AUTOMOBILE','MY_THES');
```

dbms_output.put_line('The preferred term for automobile is: '||terms);
end;

Related Topics

OUTPUT_STYLE

Preferred Term (PT) Operator in Chapter 3, "Oracle Text CONTAINS Query Operators"
This function returns the related terms of a term in the specified thesaurus.

Syntax 1: Table Result

```sql
CTX_THES.RT(restab IN OUT NOCOPY EXP_TAB,
            phrase IN VARCHAR2,
            tname  IN VARCHAR2 DEFAULT 'DEFAULT');
```

Syntax 2: String Result

```sql
CTX_THES.RT(phrase IN VARCHAR2,
            tname  IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN varchar2;
```

**restab**
Optionally, specify the name of the expansion table to store the results. This table must be of type `EXP_TAB` which the system defines as follows:

```sql
type exp_rec is record {
  xrel varchar2(12),
  xlevel number,
  xphrase varchar2(256)
};
type exp_tab is table of exp_rec index by binary_integer;
```

**See Also**: "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about `EXP_TAB`.

**phrase**
Specify a phrase to lookup in thesaurus.

**tname**
Specify the thesaurus name. If not specified, system default thesaurus is used.

**Returns**
This function returns a string of related terms in the form:

```
{rt1}|{rt2}|{rt3}| ...
```

**Example**
Consider a thesaurus `MY_THES` with the following related term definition for dog:

```
DOG
  RT WOLF
  RT HYENA
```

To look up the related terms for `dog`, execute the following code:

```sql
declare
    terms varchar2(2000);
begin
    terms := ctx_thes.rt('DOG','MY_THES');
    dbms_output.put_line('The related terms for dog are: '||terms);
end;
```
This code produces the following output:

The related terms for dog are: {dog}|{wolf}|{hyena}

Related Topics

OUTPUT_STYLE

Related Term (RT) Operator in Chapter 3, "Oracle Text CONTAINS Query Operators"
This function returns the scope note of the given phrase.

Syntax

```
CTX_THES.SN(phrase IN VARCHAR2,
    tname  IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN VARCHAR2;
```

**phrase**
Specify a phrase to lookup in thesaurus.

**tname**
Specify the thesaurus name. If not specified, system default thesaurus is used.

Returns

This function returns the scope note as a string.

Example

```
declare
    note varchar2(80);
begin
    note := ctx_thes.sn('camera','mythes');
    dbms_output.put_line('CAMERA');
    dbms_output.put_line(' SN ' || note);
end;
```

Sample output:

```
CAMERA
SN Optical cameras
```
This function returns all synonyms of a phrase as recorded in the specified thesaurus.

**Syntax 1: Table Result**

```
CTX_THESES.SYN(restab IN OUT NOCOPY EXP_TAB,
    phrase IN VARCHAR2,
    tname  IN VARCHAR2 DEFAULT 'DEFAULT');
```

**Syntax 2: String Result**

```
CTX_THESES.SYN(phrase IN VARCHAR2,
    tname  IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN VARCHAR2;
```

**restab**

Optionally, specify the name of the expansion table to store the results. This table must be of type `EXP_TAB` which the system defines as follows:

```
type exp_rec is record {
    xrel varchar2(12),
    xlevel number,
    xphrase varchar2(256)
};
type exp_tab is table of exp_rec index by binary_integer;
```

**See Also:** "CTX_THESES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about `EXP_TAB`.

**phrase**

Specify a phrase to lookup in thesaurus.

**tname**

Specify the thesaurus name. If not specified, system default thesaurus is used.

**Returns**

This function returns a string of the form:

```
{syn1}|{syn2}|{syn3} ...
```

**Example**

**String Result**

Consider a thesaurus named `ANIMALS` that has an entry for `cat` as follows:

```
CAT
    SYN KITTY
    SYN FELINE
```

To look-up the synonym for `cat` and obtain the result as a string, enter the following statements:

```
declare
    synonyms varchar2(2000);
begin
```

synonyms := ctx_thes.syn('CAT','ANIMALS');
dbms_output.put_line('the synonym expansion for CAT is: '||synonyms);
end;

This code produces the following output:
the synonym expansion for CAT is: {CAT}|{KITTY}|{FELINE}

**Table Result**
The following code looks up the synonyms for *canine* and obtains the results in a table. The contents of the table are printed to the standard output.

```sql
declare
    xtab ctx_thes.exp_tab;
begin
    ctx_thes.syn(xtab, 'canine', 'my_thesaurus');
    for i in 1..xtab.count loop
        dbms_output.put_line(lpad(' ', 2*xtab(i).xlevel) ||
            xtab(i).xrel || ' ' || xtab(i).xphrase);
    end loop;
end;
```

This code produces the following output:

<table>
<thead>
<tr>
<th>PHRASE</th>
<th>CANINE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT</td>
<td>DOG</td>
</tr>
<tr>
<td>SYN</td>
<td>PUPPY</td>
</tr>
<tr>
<td>SYN</td>
<td>MUTT</td>
</tr>
<tr>
<td>SYN</td>
<td>MONGREL</td>
</tr>
</tbody>
</table>

**Related Topics**

- **OUTPUT_STYLE**
- **SYNonym (SYN) Operator in Chapter 3, "Oracle Text CONTAINS Query Operators"**
The procedure finds and returns all top terms of a thesaurus. A top term is defined as any term which has a narrower term but has no broader terms.

This procedure differs from TT in that TT takes in a phrase and finds the top term for that phrase, but THES_TT searches the whole thesaurus and finds all top terms.

**Large Thesauri**

Because this procedure searches the whole thesaurus, it can take some time on large thesauri. Oracle recommends that you not call this often for such thesauri. Instead, your application should call this once, store the results in a separate table, and use those stored results.

**Syntax**

```sql
CTX_THES.THES_TT(restab IN OUT NOCOPY EXP_TAB,
               tname  IN VARCHAR2 DEFAULT 'DEFAULT');
```

**restab**

Specify the name of the expansion table to store the results. This table must be of type EXP_TAB which the system defines as follows:

```sql
type exp_rec is record {
   xrel  varchar2(12),
   xlevel number,
   xphrase varchar2(256)
};
type exp_tab is table of exp_rec index by binary_integer;
```

**See Also:** "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about EXP_TAB.

**tname**

Specify the thesaurus name. If not specified, system default thesaurus is used.

**Returns**

This procedure returns all top terms and stores them in restab.
For a given mono-lingual thesaurus, this function returns the foreign language equivalent of a phrase as recorded in the thesaurus.

**Syntax 1: Table Result**

```sql
CTX_THES.TR(restab IN OUT NOCOPY EXP_TAB,
            phrase IN VARCHAR2,
            lang   IN VARCHAR2 DEFAULT NULL,
            tname  IN VARCHAR2 DEFAULT 'DEFAULT')
```

**Syntax 2: String Result**

```sql
CTX_THES.TR(phrase IN VARCHAR2,
            lang   IN VARCHAR2 DEFAULT NULL,
            tname  IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN VARCHAR2;
```

**restab**
Optionally, specify the name of the expansion table to store the results. This table must be of type EXP_TAB which the system defines as follows:

```sql
type exp_rec is record (
    xrel varchar2(12),
    xlevel number,
    xphrase varchar2(256)
);
type exp_tab is table of exp_rec index by binary_integer;
```

**phrase**
Specify a phrase to lookup in thesaurus.

**lang**
Specify the foreign language. Specify 'ALL' for all translations of phrase.

**tname**
Specify the thesaurus name. If not specified, system default thesaurus is used.

**Returns**
This function returns a string of foreign terms in the form:

```
{ft1}|{ft2}|{ft3} ...
```

**Example**
Consider a thesaurus MY_THES with the following entries for cat:

---

**Note:** Foreign language translation is not part of the ISO-2788 or ANSI Z39.19 thesaural standards. The behavior of TR is specific to Oracle Text.

**See Also:** "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about EXP_TAB.
cat
  SPANISH: gato
  FRENCH: chat
  SYN lion
  SPANISH: leon

To look up the translation for cat, enter the following statements:

declare
  trans    varchar2(2000);
  span_trans varchar2(2000);
begin
  trans := ctx_thes.tr('CAT','ALL','MY_THES');
  span_trans := ctx_thes.tr('CAT','SPANISH','MY_THES')
  dbms_output.put_line('the translations for CAT are: '||trans);
  dbms_output.put_line('the Spanish translations for CAT are: '||span_trans);
end;

This codes produces the following output:

the translations for CAT are: (CAT)|(CHAT)|(GATO)
the Spanish translations for CAT are: (CAT)|(GATO)

Related Topics

OUTPUT_STYLE

Translation Term (TR) Operator in Chapter 3, "Oracle Text CONTAINS Query Operators"
For a given mono-lingual thesaurus, this function returns the foreign equivalent of a phrase, synonyms of the phrase, and foreign equivalent of the synonyms as recorded in the specified thesaurus.

Note: Foreign language translation is not part of the ISO-2788 or ANSI Z39.19 thesaural standards. The behavior of TRSYN is specific to Oracle Text.

Syntax 1: Table Result

```
CTX_THES.TRSYN(restab IN OUT NOCOPY EXP_TAB,
    phrase IN VARCHAR2,
    lang   IN VARCHAR2 DEFAULT NULL,
    tname  IN VARCHAR2 DEFAULT 'DEFAULT');
```

Syntax 2: String Result

```
CTX_THES.TRSYN(phrase IN VARCHAR2,
    lang   IN VARCHAR2 DEFAULT NULL,
    tname  IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN VARCHAR2;
```

restab
Optionally, specify the name of the expansion table to store the results. This table must be of type EXP_TAB which the system defines as follows:

```
type exp_rec is record {
    xrel varchar2(12),
    xlevel number,
    xphrase varchar2(256)
};;
type exp_tab is table of exp_rec index by binary_integer;
```

See Also: "CTX_THES Result Tables and Data Types" in Appendix A,"Oracle Text Result Tables" for more information about EXP_TAB.

phrase
Specify a phrase to lookup in thesaurus.

lang
Specify the foreign language. Specify 'ALL' for all translations of phrase.

tname
Specify the thesaurus name. If not specified, system default thesaurus is used.

Returns
This function returns a string of foreign terms in the form:

```
{ft1}|{ft2}|{ft3} ...
```

Example
Consider a thesaurus MY_THES with the following entries for cat:
cat
  SPANISH: gato
  FRENCH: chat
  SYN lion
    SPANISH: leon

To look up the translation and synonyms for cat, enter the following statements:

```sql
declare
    synonyms   varchar2(2000);
    span_syn   varchar2(2000);
begin
    synonyms := ctx_thes.trsyn('CAT','ALL','MY_THES');
    span_syn := ctx_thes.trsyn('CAT','SPANISH','MY_THES')
    dbms_output.put_line('all synonyms for CAT are: '||synonyms);
    dbms_output.put_line('the Spanish synonyms for CAT are: '||span_syn);
end;
```

This codes produces the following output:

all synonyms for CAT are: {CAT}|{CHAT}|{GATO}|{LION}|{LEON}
the Spanish synonyms for CAT are: {CAT}|{GATO}|{LION}|{LEON}

Related Topics

OUTPUT_STYLE
Translation Term Synonym (TRSYN) Operator in Chapter 3, "Oracle Text CONTAINS Query Operators"
This function returns the top term of a phrase as recorded in the specified thesaurus.

Syntax 1: Table Result

```
CTX_THES.TT(restab IN OUT NOCOPY EXP_TAB,
phrase IN VARCHAR2,
tname  IN VARCHAR2 DEFAULT 'DEFAULT');
```

Syntax 2: String Result

```
CTX_THES.TT(phrase IN VARCHAR2,
tname  IN VARCHAR2 DEFAULT 'DEFAULT')
RETURN varchar2;
```

**restab**
Optionally, specify the name of the expansion table to store the results. This table must be of type EXP_Tab which the system defines as follows:

```sql
type exp_rec is record {
  xrel varchar2(12),
  xlevel number,
  xphrase varchar2(256)
};
type exp_tab is table of exp_rec index by binary_integer;
```

**See Also:** "CTX_THES Result Tables and Data Types" in Appendix A, "Oracle Text Result Tables" for more information about EXP_TAB.

**phrase**
Specify a phrase to lookup in thesaurus.

**tname**
Specify the thesaurus name. If not specified, system default thesaurus is used.

**Returns**
This function returns the top term string in the form:

```
{tt}
```

**Example**
Consider a thesaurus MY_THES with the following broader term entries for dog:

```
DOG
   BT1 CANINE
   BT2 MAMMAL
   BT3 VERTEBRATE
   BT4 ANIMAL
```

To look up the top term for DOG, execute the following code:

```sql
declare
terms varchar2(2000);
begin
terms := ctx_thes.tt('DOG','MY_THES');
```
dbms_output.put_line('The top term for DOG is: '||terms);
end;

This code produces the following output:
The top term for dog is: {ANIMAL}

Related Topics

OUTPUT_STYLE
Top Term (TT) Operator in Chapter 3, "Oracle Text CONTAINS Query Operators"
UPDATE_TRANSLATION

Use this procedure to update an existing translation.

Syntax

```sql
CTX_THES.UPDATE_TRANSLATION(tname       in     varchar2,
phrase      in     varchar2,
language    in     varchar2,
translation in     varchar2,
new_translation in varchar2);
```

tname
Specify the name of the thesaurus, using no more than 30 characters.

phrase
Specify the phrase in the thesaurus to which to update a translation. The phrase must already exist in the thesaurus or an error is raised.

language
Specify the language of the translation, using no more than 10 characters.

translation
Specify the translated term to update. If no such translation exists, an error is raised.
You can specify NULL if there is only one translation for the phrase. An error is raised if there is more than one translation for the term in the specified language.

new_translation
Optionally, specify the new form of the translated term.

Example

The following code updates the Spanish translation for dog:

```sql
begin
    ctx_thes.update_translation('my_thes', 'dog', 'SPANISH:', 'PERRO', 'CAN');
end;
```
This chapter provides reference information on how to use the CTX_ULEXER PL/SQL package with the user-defined lexer.

CTX_ULEXER declares the following type:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WILDCARD_TAB</td>
<td>Index-by table type that you use to specify the offset of characters to be treated as wildcard characters by the user-defined lexer query procedure.</td>
</tr>
</tbody>
</table>
TYPE WILDCARD_TAB IS TABLE OF NUMBER INDEX BY BINARY_INTEGER;

Use this index-by table type to specify the offset of those characters in the query word to be treated as wildcard characters by the user-defined lexer query procedure.

Character offset information follows USC-2 codepoint semantics.
This chapter discusses the utilities shipped with Oracle Text. The following topics are included:

- **Thesaurus Loader (ctxload)**
- **Entity Extraction User Dictionary Loader (ctxload)**
- **Knowledge Base Extension Compiler (ctxkbtc)**
- **Lexical Compiler (ctxlc)**

### Thesaurus Loader (ctxload)

Use `ctxload` to import a thesaurus file into the Oracle Text thesaurus tables.

An import file is an ASCII flat file that contains entries for synonyms, broader terms, narrower terms, or related terms, which can be used to expand queries.

**See Also:** For examples of import files for thesaurus importing, see "Structure of ctxload Thesaurus Import File" on page C-3 in Appendix C, "Text Loading Examples for Oracle Text"

### Text Loading

The `ctxload` program no longer supports the loading of text columns. To load files to a text column in batch mode, Oracle recommends that you use SQL*Loader.

**See Also:** "SQL*Loader Example" on page C-1 in Appendix C, "Text Loading Examples for Oracle Text"

### ctxload Syntax

```
ctxload -user username[/password]@[sqlnet_address]
   -name object_name
   -file file_name

   [-thes]
   [-thescase y|n]
   [-thesdump]
   [-log file_name]
   [-trace]
   [-drop]
```
Mandatory Arguments

-user
Specify the user name and password of the user running ctxload.

The user name and password can be followed immediately by @sqlnet_address to permit logging on to remote databases. The value for sqlnet_address is a database connect string. If the TWO_TASK environment variable is set to a remote database, then you do not need to specify a value for sqlnet_address to connect to the database.

-name object_name
When you use ctxload to import a thesaurus, use object_name to specify the name of the thesaurus to be imported.

Use object_name to identify the thesaurus in queries that use thesaurus operators.

Note: Thesaurus name must be unique. If the name specified for the thesaurus is identical to an existing thesaurus, then ctxload returns an error and does not overwrite the existing thesaurus.

-file file_name
When ctxload is used to import a thesaurus, use file_name to specify the name of the import file that contains the thesaurus entries.

When ctxload is used to export a thesaurus, use file_name to specify the name of the export file created by ctxload.

Note: If the name specified for the thesaurus dump file is identical to an existing file, then ctxload overwrites the existing file.

Optional Arguments

-thes
Import a thesaurus. Specify the source file with the -file argument. Specify the name of the thesaurus to be imported with -name.

-thescase y | n
Specify y to create a case-sensitive thesaurus with the name specified by -name and populate the thesaurus with entries from the thesaurus import file specified by -file. If -thescase is y (the thesaurus is case-sensitive), ctxload enters the terms in the thesaurus exactly as they appear in the import file.

The default for -thescase is n (case-insensitive thesaurus).

Note: -thescase is valid for use only with the -thes argument.

-thesdump
Export a thesaurus. Specify the name of the thesaurus to be exported with the -name argument. Specify the destination file with the -file argument.
Specify the name of the log file to which `ctxload` writes any national-language supported (Globalization Support) messages generated during processing. If you do not specify a log file name, the messages appear on the standard output.

Enables SQL statement tracing using `ALTER SESSION SET SQL_TRACE TRUE`. This command captures all processed SQL statements in a trace file, which can be used for debugging. The location of the trace file is operating-system dependent and can be modified using the `USER_DUMP_DEST` initialization parameter.

See Also: For more information about SQL trace and the `USER_DUMP_DEST` initialization parameter, see Oracle Database Administrator’s Guide

cxtload Examples

This section provides examples for some of the operations that `ctxload` can perform.

See Also: For more document loading examples, see Appendix C, "Text Loading Examples for Oracle Text"

Thesaurus Import Example

The following example imports a thesaurus named `tech_doc` from an import file named `tech_thesaurus.txt`:

```
ctxload -user jsmith/123abc -thes -name tech_doc -file tech_thesaurus.txt
```

Thesaurus Export Example

The following example dumps the contents of a thesaurus named `tech_doc` into a file named `tech_thesaurus.out`:

```
ctxload -user jsmith/123abc -thesdump -name tech_doc -file tech_thesaurus.out
```

Entity Extraction User Dictionary Loader (ctxload)

Use `ctxload` to import an entity extraction user dictionary into Oracle Text tables.

An import file is an XML flat file containing entries for entities, with their associated types and alternate forms.

cxtload Syntax

```
ctxload -user username[/password][@sqlnet_address]
   -extract
   -name entity extraction policy name
   -file user-dictionary file name
   [-drop] to drop a user-dictionary from a policy
```

Mandatory Arguments

-`-user`

Specify the user name and password of the user running `ctxload`.

The user name and password can be followed immediately by `@sqlnet_address` to permit logging on to remote databases. The value for `sqlnet_address` is a database name.
connect string. If the \texttt{TWO\_TASK} environment variable is set to a remote database, then you do not need to specify a value for \texttt{sqlnet\_address} to connect to the database.

\textbf{-name entity extraction policy name}

When you use \texttt{ctxload} to import an entity extraction dictionary, use \texttt{object\_name} to specify the entity extraction policy to associate the dictionary with. An entity extraction policy can have only one user dictionary.

\textbf{-file user-dictionary file name}

Use \texttt{file} to specify the name of the XML file containing the user dictionary.

\textbf{-drop}

Drop the user dictionary currently associated with an entity extraction policy.

\section*{Considerations When Creating a User Dictionary}

The following are some considerations for when creating a user dictionary:

\begin{itemize}
  \item Entity mentions are case-sensitive. They cannot contain any null characters.
  \item Entity type names are case-insensitive. They cannot contain any null or comma characters.
  \item Customers will be able to assign two or more entity types to a single entity mention. For example, the entity "Washington" could be assigned the type "CITY" and also the type "STATE".
  \item The content of a user's dictionary is invisible to other users.
  \item The maximum byte length of an entity mention is 512 bytes by the server-side database character set.
  \item The maximum byte length of an entity type name is 30 bytes by the server-side database character set.
  \item User-defined entity type names must start with the letter "x".
\end{itemize}

\section*{XML Schema}

The entity extraction dictionary follows this XML schema:

\begin{verbatim}
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">

  <xsd:element name="dictionary">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="entities" type="entityType" maxOccurs="unbounded"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>

  <xsd:complexType name="entityType">
    <xsd:sequence>
      <xsd:element name="entity" type="entType" maxOccurs="unbounded"/>
    </xsd:sequence>
    <xsd:attribute name="language" type="xsd:string"/>
  </xsd:complexType>

  <xsd:complexType name="entType">
    <xsd:sequence>
      <xsd:element name="value" type="xsd:string"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:schema>
\end{verbatim}
The following tables illustrate some aspects of the XML schema for the entity extraction dictionary.

<table>
<thead>
<tr>
<th>Element Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dictionary</td>
<td>Collection of entities</td>
</tr>
<tr>
<td>entities</td>
<td>Collection of entities per language</td>
</tr>
<tr>
<td>entity</td>
<td>Each entity</td>
</tr>
<tr>
<td>value</td>
<td>Entity mention</td>
</tr>
<tr>
<td>type</td>
<td>Entity type</td>
</tr>
<tr>
<td>alternate</td>
<td>Alternate form of entity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Element Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>language</td>
<td>Entities</td>
<td>Language name of each entity in entities</td>
</tr>
</tbody>
</table>

**ctxload Example**

The following is an example of an entity extraction user dictionary file that can be loaded using `ctxload`:

```xml
<?xml version='1.0' encoding='utf-8'?>
<dictionary>
  <entities>
    <entity>
      <value>New York</value>
      <type>city</type>
    </entity>
  </entities>
  <entities language="german">
    <entity>
      <value>Deutschland</value>
      <type>country</type>
    </entity>
  </entities>
  <entities language="english">
    <entity>
      <value>Astra</value>
      <type>person</type>
      <type>organization</type>
    </entity>
    <entity>
      <value>George W. Bush</value>
      <type>person</type>
      <alternate>G. W. Bush</alternate>
      <alternate>G. Bush</alternate>
    </entity>
  </entities>
</dictionary>
```
Knowledge Base Extension Compiler (ctxkbtc)

The knowledge base is the information source that Oracle Text uses to perform theme analysis, such as theme indexing, processing ABOUT queries, and to document theme extraction with the CTX_DOC package. A knowledge base is supplied for English and French.

With the ctxkbtc compiler, you can:

- Extend your knowledge base by compiling one or more thesauri with the Oracle Text knowledge base. The extended information can be application-specific terms and relationships. During theme analysis, the extended portion of the knowledge base overrides any terms and relationships in the knowledge base where there is overlap.
- Create a new user-defined knowledge base by compiling one or more thesauri. In languages other than English and French, this feature can be used to create a language-specific knowledge base.

**Note:** Only CTXSYS can extend the knowledge base.

**See Also:** For more information about the knowledge base packaged with Oracle Text, see [http://www.oracle.com/technology/products/text/](http://www.oracle.com/technology/products/text/)

For more information about the ABOUT operator, see ABOUT operator in Chapter 3, "Oracle Text CONTAINS Query Operators"

For more information about document services, see Chapter 8, "CTX_DOC Package"

Knowledge Base Character Set

Knowledge bases can be in any single-byte character set. Supplied knowledge bases are in WE8ISO8859P1. You can store an extended knowledge base in another character set such as US7ASCII.

ctxkbtc Syntax

ctxkbtc -user uname/passwd
[-name thesname1 [thesname2 ... thesname16]]
[-revert]
[-stoplist stoplistname]
[-verbose]
[-log filename]

-**user**
Specify the user name and password for the administrator creating an extended knowledge base. This user must have write permission to the ORACLE_HOME directory.

-**name thesname1 [thesname2 ... thesname16]**
Specify the names of the thesauri (up to 16) to be compiled with the knowledge base to create the extended knowledge base. The thesauri you specify must already be loaded with ctxload with the "-thescase Y" option

-**revert**
Reverts the extended knowledge base to the default knowledge base provided by Oracle Text.
-stoplist stoplistname
Specify the name of the stoplist. Stopwords in the stoplist are added to the knowledge base as useless words that are prevented from becoming themes or contributing to themes. Add stopthemes after running this command using CTX_DLL.ADD_STOPTHEME.

-verbose
Displays all warnings and messages, including non-Globalization Support messages, to the standard output.

-log
Specify the log file for storing all messages. When you specify a log file, no messages are reported to standard out.

ctxkbtc Usage Notes
- Before running ctxkbtc, you must set the NLS_LANG environment variable to match the database character set.
- The user issuing ctxkbtc must have write permission to the ORACLE_HOME, because the program writes files to this directory.
- Before being compiled, each thesaurus must be loaded into Oracle Text case sensitive with the "-thescase Y" option in ctxload.
- Running ctxkbtc twice removes the previous extension.

cctxkbtc Limitations
The ctxkbtc program has the following limitations:
- When upgrading or downgrading your database to a different release, for theme indexing and related features to work correctly, Oracle recommends that you recompile your extended knowledge base in the new environment.
- Before extending the knowledge base, you must terminate all server processes that have invoked any knowledge base-related Text functions during their lifetime.
- There can be only one user extension for each language for each installation. Because a user extension affects all users at the installation, only the CTXSYS user can extend the knowledge base.
- In an Oracle RAC environment, the ORACLE_HOME can either be shared between multiple nodes, or each node can have its own ORACLE_HOME. The following requirements apply:
  - Before using any knowledge base-dependent functionality in any of the Oracle RAC nodes, ctxkbtc must be run in every ORACLE_HOME in the Oracle RAC environment.
  - When using ctxkbtc, the exact same input thesaurus content must be used in every ORACLE_HOME in the Oracle RAC environment.

cctxkbtc Constraints on Thesaurus Terms
Terms are case sensitive. If a thesaurus has a term in uppercase, for example, the same term present in lowercase form in a document will not be recognized.
The maximum length of a term is 80 characters.
Disambiguated homographs are not supported.
ctxkbtc Constraints on Thesaurus Relations

The following constraints apply to thesaurus relations:

- BTG and BTP are the same as BT. NTG and NTP are the same as NT.
- Only preferred terms can have a BT, NTs or RTs.
- If a term has no USE relation, it will be treated as its own preferred term.
- If a set of terms are related by SYN relations, only one of them may be a preferred term.
- An existing category cannot be made a top term.
- There can be no cycles in BT and NT relations.
- A term can have at most one preferred term and at most one BT. A term may have any number of NTs.
- An RT of a term cannot be an ancestor or descendant of the term. A preferred term may have any number of RTs up to a maximum of 32.
- The maximum height of a tree is 16 including the top term level.
- When multiple thesauri are being compiled, a top term in one thesaurus should not have a broader term in another thesaurus.

---

Note: The thesaurus compiler tolerates some violations of the preceding rules. For example, if a term has multiple BTs, then the compiler ignores all but the last one it encounters.

Similarly, BTs between existing knowledge base categories result only in a warning message.

Oracle recommends that you do not set up extended storage bases with violations. Using extended storage bases containing violations can produce undesired results.

---

Extending the Knowledge Base

Extend the supplied knowledge base by compiling one or more thesauri with the Oracle Text knowledge base. The extended information can be application-specific terms and relationships. During theme analysis, the extended portion of the knowledge base overrides any terms and relationships in the knowledge base where there is overlap.

When extending the knowledge base, Oracle recommends that new terms be linked to one of the categories in the knowledge base for best results in theme proving when appropriate.

See Also: For complete description of the supplied knowledge base, see

http://www.oracle.com/technology/products/text/
Example for Extending the Knowledge Base

You purchase a medical thesaurus `medthes` containing a hierarchy of medical terms. The four top terms in the thesaurus are the following:

- Anesthesia and Analgesia
- Anti-Allergic and Respiratory System Agents
- Anti-Inflammatory Agents, Antirheumatic Agents, and Inflammation Mediators
- Antineoplastic and Immunosuppressive Agents

To link these terms to the existing knowledge base, add the following entries to the medical thesaurus to map the new terms to the existing `health and medicine` branch:

```
health and medicine
   NT Anesthesia and Analgesia
   NT Anti-Allergic and Respiratory System Agents
   NT Anti-Inflammatory Agents, Antirheumatic Agents, and Inflammation Mediators
   NT Antineoplastic and Immunosuppressive Agents
```

Set your globalization support language environment variable to match the database character set. For example, if your database character set is `WE8ISO8859P1` and you are using American English, set your `NLS_LANG` as follows:

```
setenv NLS_LANG AMERICAN_AMERICA.WE8ISO8859P1
```

Assuming the medical thesaurus is in a file called `med.thes`, load the thesaurus as `medthes` with `ctxload` as follows:

```
ctxload -thes -thescase y -name medthes -file med.thes -user ctxsys/ctxsys
```

To link the loaded thesaurus `medthes` to the knowledge base, use `ctxkbtc` as follows:

```
ctxkbtc -user ctxsys/ctxsys -name medthes
```

Adding a Language-Specific Knowledge Base

Extend theme functionality to languages other than English or French by loading your own knowledge base for any single-byte whitespace delimited language, including Spanish.

Theme functionality includes theme indexing, `ABOUT` queries, theme highlighting, and the generation of themes, gists, and theme summaries with the `CTX_DOC` PL/SQL package.

Extend theme functionality by adding a user-defined knowledge base. For example, you can create a Spanish knowledge base from a Spanish thesaurus.

To load your language-specific knowledge base, follow these steps:

1. Load your custom thesaurus using `ctxload`.
2. Set `NLS_LANG` so that the language portion is the target language. The charset portion must be a single-byte character set.
3. Compile the loaded thesaurus using `ctxkbtc`:

   ```
   ctxkbtc -user ctxsys/ctxsys -name my_lang_thes
   ```

This command compiles your language-specific knowledge base from the loaded thesaurus. To use this knowledge base for theme analysis during indexing and `ABOUT`
queries, specify the NLS_LANG language as the THEME_LANGUAGE attribute value for the BASIC_LEXER preference.

Limitations for Adding a Knowledge Base

The following limitations hold for adding knowledge bases:

- Oracle Text supplies knowledge bases in English and French only. You must provide your own thesaurus for any other language.

- You can only add knowledge bases for languages with single-byte character sets. You cannot create a knowledge base for languages which can be expressed only in multibyte character sets. If the database is a multibyte universal character set, such as UTF-8, the NLS_LANG parameter must still be set to a compatible single-byte character set when compiling the thesaurus.

- Adding a knowledge base works best for whitespace delimited languages.

- You can have at most one knowledge base for each globalization support language.

- Obtaining hierarchical query feedback information such as broader terms, narrower terms and related terms does not work in languages other than English and French. In other languages, the knowledge bases are derived entirely from your thesauri. In such cases, Oracle recommends that you obtain hierarchical information directly from your thesauri.

Order of Precedence for Multiple Thesauri

When multiple thesauri are to be compiled, precedence is determined by the order in which thesauri are listed in the arguments to the compiler, assumed to be most preferred first. A user-defined thesaurus always has precedence over the built-in knowledge base.

Size Limits for Extended Knowledge Base

The following table lists the size limits associated with creating and compiling an extended knowledge base.

<table>
<thead>
<tr>
<th>Description of Parameter</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of RTs (from + to) for each term</td>
<td>32</td>
</tr>
<tr>
<td>Number of terms for each single hierarchy (for example, all narrower terms for a given top term)</td>
<td>64000</td>
</tr>
<tr>
<td>Number of new terms in an extended knowledge base</td>
<td>1 million</td>
</tr>
<tr>
<td>Number of separate thesauri that can be compiled into a user extension to the KB</td>
<td>16</td>
</tr>
</tbody>
</table>

Lexical Compiler (ctxlc)

The Lexical Compiler (ctxlc) is a command-line utility that enables you to create your own Chinese and Japanese lexicons (dictionaries). Such a lexicon may either be generated from a user-supplied word list or from the merging of a word list with the system lexicon for that language.
ctxlc creates the new lexicon in your current directory. The new lexicon consists of three files, drol.dat, drolk.dat, and droli.dat. To change your system lexicon for Japanese or Chinese, overwrite the system lexicon with these files.

The Lexical Compiler can also generate wordlists from the system lexicons for Japanese and Chinese, enabling you to see their contents. These word lists go to the standard output and thus can be redirected into a file of your choice.

After overwriting the system lexicon, you need to re-create your indexes before querying them.

**Syntax of ctxlc**

ctxlc has the following syntax:

```
ctxlc -ja | -zh [ -n ] -ics character_set -i input_file
ctxlc -ja | -zh -ocs character_set [ > output_file ]
```

**Mandatory Arguments**

- `ja` | `zh`
  Specify the language of the lexicon to modify or create. `-ja` indicates the Japanese lexicon; `-zh` indicates the Chinese lexicon.

- `-ics character_set`
  Specify the character set of the input file denoted by `-i input_file`. `input_file` is the list of words, one word to a line, to use in creating the new lexicon.

- `i input_file`
  Specify the file containing words to use in creating a new lexicon.

- `ocs character_set`
  Specify the character set of the text file to be output.

**Optional Arguments**

- `-n`
  Specify `-n` to create a new lexicon that consists only of user-supplied words taken from `input_file`. If `-n` is not specified, then the new lexicon consists of a merge of the system lexicon with `input_file`. Also, when `-n` is not selected, a text file called `drolt.dat`, is created in the current directory to enable you to inspect the contents of the merged lexicon without having to enter another `ctxlc` command.

**Performance Considerations**

You can add up to 1,000,000 new words to a lexicon. However, creating a very large lexicon can reduce performance in indexing and querying. Performance is best when the lexicon character set is UTF-8. There is no performance impact on the Chinese or Japanese V-gram lexers, as they do not use lexicons.

**ctxlc Usage Notes**

Oracle recommends the following practices with regard to `ctxlc`:

- Save your plain text dictionary file in your environment for emergency use.
When upgrading or downgrading your database to a different release, recompile your plain text dictionary file in the new environment so that the user lexicon will work correctly.

Example

In this example, you create a new Japanese lexicon from the file jadict.txt, a word list that uses the JA16EUC character set. Because you are not specifying -n, the new lexicon is the result of merging jadict.txt with the system Japanese lexicon. Then replace the existing Japanese lexicon with the new, merged one.

% ctxlc -ja -ics JA16EUC -i jadict.txt

This creates new files in the current directory:

% ls
drold.dat
drolk.dat
droli.dat
drolt.dat

The system lexicon files for Japanese and Chinese are named droldxx.dat, drolkxx.dat, and drolix.dat, where xx is either JA (for Japanese) or ZH (for Chinese). Rename the three new files and copy them to the directory containing the system Japanese lexicon.

% mv drold.dat droldJA.dat
% mv drolk.dat drolkJA.dat
% mv droli.dat droliJA.dat
% cp *dat $ORACLE_HOME/ctx/data/jalx

This replaces the system Japanese lexicon with one that is a merge of the old system lexicon and your wordlist from jadict.txt.

You can also use ctxlc to get a dump of a system lexicon. This example dumps the Chinese lexicon to a file called new_chinese_dict.txt in the current directory:

% ctxlc -zh -ocs UTF8 > new_chinese_dict.txt

This creates a file, new_japanese.dict.txt, using the UTF8 character set, in the current directory.
This chapter describes various ways that Oracle Text handles alternative spelling of words. It also documents the alternative spelling conventions that Oracle Text uses in the German, Danish, and Swedish languages.

The following topics are covered:

- Overview of Alternative Spelling Features
- Overriding Alternative Spelling Features
- Alternative Spelling Conventions

Overview of Alternative Spelling Features

Some languages have alternative spelling forms for certain words. For example, the German word \textit{Schoen} can also be spelled as \textit{Schon}.

The form of a word is either \textit{original} or \textit{normalized}. The original form of the word is how it appears in the source document. The normalized form is how it is transformed, if it is transformed at all. Depending on the word being indexed and which system preferences are in effect (these are discussed in this chapter), the normalized form of a word may be the same as the original form. Also, the normalized form may comprise more than one spelling. For example, the normalized form of \textit{Schoen} is both \textit{Schoen} and \textit{Schon}.

Oracle Text handles indexing of alternative word forms in the following ways:

- Alternate Spelling—indexing of alternative forms is enabled
- Base-Letter Conversion—accented letters are transformed into non-accented representations
- New German Spelling—reformed German spelling is accepted

Enable these features by specifying the appropriate attribute to the \texttt{BASIC_LEXER}. For instance, enable alternate spelling by specifying either \texttt{GERMAN}, \texttt{DANISH}, or \texttt{SWEDISH} for the \texttt{ALTERNATE_SPELLING} attribute. As an example, here is how to enable alternate spelling in German:

\begin{verbatim}
begin
  ctx_ddl.create_preference('GERMAN_LEX', 'BASIC_LEXER');
  ctx_ddl.set_attribute('GERMAN_LEX', 'ALTERNATE_SPELLING', 'GERMAN');
end;
\end{verbatim}

To disable alternate spelling, use the \texttt{CTX_DDL.UNSET_ATTRIBUTE} procedure as follows:

\begin{verbatim}
begin
\end{verbatim}
ctx_ddl.unset_attribute('GERMAN_LEX', 'ALTERNATE_SPELLING'); end;

Oracle Text converts query terms to their normalized forms before lookup. As a result, users can query words with either spelling. If Schoen has been indexed as both Schoen and Schön, a query with Schön returns documents containing either form.

**Alternate Spelling**

When Swedish, German, or Danish has more than one way of spelling a word, Oracle Text normally indexes the word in its original form; that is, as it appears in the source document.

When Alternate Spelling is enabled, Oracle Text indexes words in their normalized form. So, for example, Schoen is indexed both as Schoen and as Schön, and a query on Schoen will return documents containing either spelling. (The same is true of a query on Schön.)

To enable Alternate Spelling, set the `BASIC_LEXER` attribute `ALTERNATE_SPELLING` to `GERMAN`, `DANISH`, or `SWEDISH`. See "BASIC_LEXER" on page 2-39 for more information.

**Base-Letter Conversion**

Besides alternative spelling, Oracle Text also handles base-letter conversions. With base-letter conversions enabled, letters with umlauts, acute accents, cedillas, and the like are converted to their basic forms for indexing, so fiancé is indexed both as fiancé and as fiance, and a query of fiancé returns documents containing either form.

To enable base-letter conversions, set the `BASIC_LEXER` attribute `BASE_LETTER` to `YES`. See "BASIC_LEXER" on page 2-39 for more information.

When Alternate Spelling is also enabled, Base-Letter Conversion may need to be overridden to prevent unexpected results. See "Overriding Base-Letter Transformations with Alternate Spelling" on page 16-3 for more information.

**Generic Versus Language-Specific Base-Letter Conversions**

The `BASE_LETTER_TYPE` attribute affects the way base-letter conversions take place. It has two possible values: `GENERIC` or `SPECIFIC`.

The `GENERIC` value is the default and specifies that base letter transformation uses one transformation table that applies to all languages.

The `SPECIFIC` value means that a base-letter transformation that has been specifically defined for your language will be used. This enables you to use accent-sensitive searches for words in your own language, while ignoring accents that are from other languages.

For example, both the `GENERIC` and the Spanish `SPECIFIC` tables will transform é into e. However, they treat the letter ñ distinctly. The `GENERIC` table treats ñ as an n with an accent (actually, a tilde), and so transforms ñ to n. The Spanish `SPECIFIC` table treats ñ as a separate letter of the alphabet, and thus does not transform it.

**New German Spelling**

In 1996, new spelling rules for German were approved by representatives from all German-speaking countries. For example, under the spelling reforms, Potential becomes Potenzial, Schiffahrt becomes Schiffahrt, and schneuzen becomes schnäuzen.
When the `BASIC_LEXER` attribute `NEW_GERMAN_SPELLING` is set to `YES`, then a `CONTAINS` query on a German word that has both new and traditional forms will return documents matching both forms. For example, a query on `Potential` returns documents containing both `Potential` and `Potenzial`. The default setting is `NO`.

---

**Note:** Under reformed German spelling, many words traditionally spelled as one word, such as `soviel`, are now spelled as two (`so viel`). Currently, Oracle Text does not make these conversions, nor conversions from two words to one (for example, `weh tun` to `wehtun`).

The case of the transformed word is determined from the first two characters of the word in the source document; that is, `schiffahrt` becomes `schiiffahrt`, `Schiffahrt` becomes `Schiiffahrt`, and `SCHIFFAHRT` becomes `SCHIFFFAHRT`.

As many new German spellings include hyphens, it is recommended that users choosing `NEW_GERMAN_SPELLING` define hyphens as `printjoints`.

See "BASIC_LEXER" on page 2-39 for more information on setting this attribute.

### Overriding Alternative Spelling Features

Even when alternative spelling features have been specified by lexer preference, it is possible to override them. Overriding takes the following form:

- Overriding of base-letter conversion when Alternate Spelling is used, to prevent characters with alternate spelling forms, such as `ü`, `ö`, and `ä`, from also being transformed to the base letter forms.

### Overriding Base-Letter Transformations with Alternate Spelling

Transformations caused by turning on `alternate_spelling` are performed before those of `base_letter`, which can sometimes cause unexpected results when both are enabled.

When Alternate Spelling is enabled, Oracle Text converts two-letter forms to single-letter forms (for example, `ue` to `ü`), so that words can be searched in both their base and alternate forms. Therefore, with Alternate Spelling enabled, a search for `Schoen` will return documents with both `Schoen` and `Schön`.

However, when Base-letter Transformation is also enabled, the `ü` in `Schlüssel` is transformed into a `u`, producing the non-existent word (in German, anyway) `Schlussel`, and the word is indexed in all three forms.

To prevent this secondary conversion, set the `OVERWRITE_BASE_LETTER` attribute to `TRUE`.

`OVERWRITE_BASE_LETTER` only affects letters with umlauts; accented letters, for example, are still transformed into their base forms.

For more on `BASE_LETTER`, see "Base-Letter Conversion" on page 16-2.

### Alternative Spelling Conventions

The following sections show the alternative spelling substitutions used by Oracle Text.
German Alternate Spelling Conventions

The German alphabet is the English alphabet plus the additional characters: ä ö ü ß. Table 16–1 lists the alternate spelling conventions Oracle Text uses for these characters.

<table>
<thead>
<tr>
<th>Character</th>
<th>Alternate Spelling Substitution</th>
</tr>
</thead>
<tbody>
<tr>
<td>ä</td>
<td>ae</td>
</tr>
<tr>
<td>ü</td>
<td>ue</td>
</tr>
<tr>
<td>ö</td>
<td>oe</td>
</tr>
<tr>
<td>Ä</td>
<td>AE</td>
</tr>
<tr>
<td>Ü</td>
<td>UE</td>
</tr>
<tr>
<td>Ö</td>
<td>OE</td>
</tr>
<tr>
<td>ß</td>
<td>ss</td>
</tr>
</tbody>
</table>

Danish Alternate Spelling Conventions

The Danish alphabet is the Latin alphabet without the w, plus the special characters: ø æ å. Table 16–2 lists the alternate spelling conventions Oracle Text uses for these characters.

<table>
<thead>
<tr>
<th>Character</th>
<th>Alternate Spelling Substitution</th>
</tr>
</thead>
<tbody>
<tr>
<td>æ</td>
<td>ae</td>
</tr>
<tr>
<td>ø</td>
<td>oe</td>
</tr>
<tr>
<td>å</td>
<td>aa</td>
</tr>
<tr>
<td>Æ</td>
<td>AE</td>
</tr>
<tr>
<td>Ø</td>
<td>OE</td>
</tr>
<tr>
<td>Å</td>
<td>AA</td>
</tr>
</tbody>
</table>

Swedish Alternate Spelling Conventions

The Swedish alphabet is the English alphabet without the w, plus the additional characters: å ä ö. Table 16–3 lists the alternate spelling conventions Oracle Text uses for these characters.

<table>
<thead>
<tr>
<th>Character</th>
<th>Alternate Spelling Convention</th>
</tr>
</thead>
<tbody>
<tr>
<td>å</td>
<td>ae</td>
</tr>
<tr>
<td>å</td>
<td>aa</td>
</tr>
<tr>
<td>ö</td>
<td>oe</td>
</tr>
<tr>
<td>Ä</td>
<td>AE</td>
</tr>
<tr>
<td>Å</td>
<td>AA</td>
</tr>
<tr>
<td>Ö</td>
<td>OE</td>
</tr>
</tbody>
</table>
This appendix describes the structure of the result tables used to store the output generated by the procedures in the CTX_QUERY, CTX_DOC, and CTX_THES packages.

The following topics are discussed in this appendix:

- CTX_QUERY Result Tables
- CTX_DOC Result Tables
- CTX_THES Result Tables and Data Types

**CTX_QUERY Result Tables**

For the CTX_QUERY procedures that return results, tables for storing the results must be created before the procedure is called. The tables can be named anything, but must include columns with specific names and data types.

This section describes the following types of result tables, and their required columns:

- EXPLAIN Table
- HFEEDBACK Table

**EXPLAIN Table**

Table A–1 describes the structure of the table to which CTX_QUERY.EXPLAIN writes its results.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Datatype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPLAIN_ID</td>
<td>VARCHAR2(30)</td>
<td>The value of the explain_id argument specified in the FEEDBACK call.</td>
</tr>
<tr>
<td>ID</td>
<td>NUMBER</td>
<td>A number assigned to each node in the query execution tree. The root operation node has ID =1. The nodes are numbered in a top-down, left-first manner as they appear in the parse tree.</td>
</tr>
<tr>
<td>PARENT_ID</td>
<td>NUMBER</td>
<td>The ID of the execution step that operates on the output of the ID step. Graphically, this is the parent node in the query execution tree. The root operation node (ID =1) has PARENT_ID = 0.</td>
</tr>
<tr>
<td>OPERATION</td>
<td>VARCHAR2(30)</td>
<td>Name of the internal operation performed. Refer to Table A–2 for possible values.</td>
</tr>
</tbody>
</table>
Table A–2  EXPLAIN Table OPERATION Column

<table>
<thead>
<tr>
<th>Operation Value</th>
<th>Query Operator</th>
<th>Equivalent Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABOUT</td>
<td>ABOUT</td>
<td>(none)</td>
</tr>
<tr>
<td>ACCUMULATE</td>
<td>ACCUM</td>
<td>,</td>
</tr>
<tr>
<td>AND</td>
<td>AND</td>
<td>&amp;</td>
</tr>
<tr>
<td>COMPOSITE</td>
<td>(none)</td>
<td>(none)</td>
</tr>
<tr>
<td>EQUIVALENCE</td>
<td>EQUIV</td>
<td>=</td>
</tr>
<tr>
<td>MINUS</td>
<td>MINUS</td>
<td>-</td>
</tr>
<tr>
<td>NEAR</td>
<td>NEAR</td>
<td>;</td>
</tr>
<tr>
<td>NOT</td>
<td>NOT</td>
<td>~</td>
</tr>
<tr>
<td>NO_HITS</td>
<td>(no hits will result from this query)</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>OR</td>
<td></td>
</tr>
<tr>
<td>PHRASE</td>
<td>(a phrase term)</td>
<td></td>
</tr>
<tr>
<td>SECTION</td>
<td>(section)</td>
<td></td>
</tr>
<tr>
<td>THRESHOLD</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>WEIGHT</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>WITHIN</td>
<td>within</td>
<td>(none)</td>
</tr>
<tr>
<td>WORD</td>
<td>(a single term)</td>
<td></td>
</tr>
</tbody>
</table>

OPTIONS Column Values

Table A–3 list the possible values for the OPTIONS column of the EXPLAIN table.

Table A–3  EXPLAIN Table OPTIONS Column

<table>
<thead>
<tr>
<th>Options Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>($)</td>
<td>Stem</td>
</tr>
<tr>
<td>(?)</td>
<td>Fuzzy</td>
</tr>
</tbody>
</table>
HFEEDBACK Table

Table A–4 describes the table to which CTX_QUERY.HFEEDBACK writes its results.

Table A–4  HFEEDBACK Results Table

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Datatype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEEDBACK_ID</td>
<td>VARCHAR2(30)</td>
<td>The value of the feedback_id argument specified in the HFEEDBACK call.</td>
</tr>
<tr>
<td>ID</td>
<td>NUMBER</td>
<td>A number assigned to each node in the query execution tree. The root operation node has ID =1. The nodes are numbered in a top-down, left-first manner as they appear in the parse tree.</td>
</tr>
<tr>
<td>PARENT_ID</td>
<td>NUMBER</td>
<td>The ID of the execution step that operates on the output of the ID step. Graphically, this is the parent node in the query execution tree. The root operation node (ID =1) has PARENT_ID = 0.</td>
</tr>
<tr>
<td>OPERATION</td>
<td>VARCHAR2(30)</td>
<td>Name of the internal operation performed. Refer to Table A–5 for possible values.</td>
</tr>
<tr>
<td>OPTIONS</td>
<td>VARCHAR2(30)</td>
<td>Characters that describe a variation on the operation described in the OPERATION column. When an OPERATION has more than one OPTIONS associated with it, OPTIONS values are concatenated in the order of processing. See Table A–6 for possible values.</td>
</tr>
<tr>
<td>OBJECT_NAME</td>
<td>VARCHAR2(80)</td>
<td>Section name, wildcard term, weight, threshold value or term to lookup in the index.</td>
</tr>
<tr>
<td>POSITION</td>
<td>NUMBER</td>
<td>The order of processing for nodes that all have the same PARENT_ID. The positions are numbered in ascending order starting at 1.</td>
</tr>
<tr>
<td>BT_FEEDBACK</td>
<td>CTX_FEEDBACK_TYPE</td>
<td>Stores broader feedback terms. See Table A–7.</td>
</tr>
<tr>
<td>PT_FEEDBACK</td>
<td>CTX_FEEDBACK_TYPE</td>
<td>Stores related feedback terms. See Table A–7.</td>
</tr>
<tr>
<td>NT_FEEDBACK</td>
<td>CTX_FEEDBACK_TYPE</td>
<td>Stores narrower feedback terms. See Table A–7.</td>
</tr>
</tbody>
</table>

Options Value  Description

(1) Soundex

(T) Order for ordered Near.

(F) Order for unordered Near.

(n) A number associated with the max_span parameter for the Near operator.

[9] Indicates that index_stems is set and query is using token_type 9.
Operation Column Values
Table A–5 shows the possible values for the OPERATION column of the HFEEDBACK table.

Table A–5  HFEEDBACK Results Table OPERATION Column

<table>
<thead>
<tr>
<th>Operation Value</th>
<th>Query Operator</th>
<th>Equivalent Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABOUT</td>
<td>ABOUT</td>
<td>(none)</td>
</tr>
<tr>
<td>ACCUMULATE</td>
<td>ACCUM</td>
<td>,</td>
</tr>
<tr>
<td>AND</td>
<td>AND</td>
<td>&amp;</td>
</tr>
<tr>
<td>EQUIVALENCE</td>
<td>EQUIV</td>
<td>=</td>
</tr>
<tr>
<td>MINUS</td>
<td>MINUS</td>
<td>–</td>
</tr>
<tr>
<td>NEAR</td>
<td>NEAR</td>
<td>;</td>
</tr>
<tr>
<td>NOT</td>
<td>NOT</td>
<td>~</td>
</tr>
<tr>
<td>OR</td>
<td>OR</td>
<td></td>
</tr>
<tr>
<td>SECTION</td>
<td>(section)</td>
<td></td>
</tr>
<tr>
<td>TEXT</td>
<td>word or phrase of a text query</td>
<td></td>
</tr>
<tr>
<td>THEME</td>
<td>word or phrase of an ABOUT query</td>
<td></td>
</tr>
<tr>
<td>THRESHOLD</td>
<td>&gt;</td>
<td>&gt;</td>
</tr>
<tr>
<td>WEIGHT</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>WITHIN</td>
<td>within</td>
<td>(none)</td>
</tr>
</tbody>
</table>

OPTIONS Column Values
Table A–6 list the values for the OPTIONS column of the HFEEDBACK table.

Table A–6  HFEEDBACK Results Table OPTIONS Column

<table>
<thead>
<tr>
<th>Options Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T)</td>
<td>Order for ordered Near.</td>
</tr>
<tr>
<td>(F)</td>
<td>Order for unordered Near.</td>
</tr>
<tr>
<td>(n)</td>
<td>A number associated with the max_span parameter for the Near operator.</td>
</tr>
</tbody>
</table>

CTX_FEEDBACK_TYPE
The CTX_FEEDBACK_TYPE is a nested table of objects. This datatype is pre-defined in the CTXSYS schema. Use this type to define the columns BT_FEEDBACK, RT_FEEDBACK, and NT_FEEDBACK.

The nested table CTX_FEEDBACK_TYPE holds objects of type CTX_FEEDBACK_ITEM_TYPE, which is also pre-defined in the CTXSYS schema. This object is defined with three members and one method as follows:

Table A–7  CTX_FEEDBACK_ITEM_TYPE

<table>
<thead>
<tr>
<th>CTX_FEEDBACK_ITEM_TYPE Members and Methods</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>text</td>
<td>NUMBER</td>
<td>Feedback term.</td>
</tr>
<tr>
<td>cardinality</td>
<td>NUMBER</td>
<td>(reserved for future use.)</td>
</tr>
</tbody>
</table>
The SQL code that defines these objects is as follows:

```sql
CREATE OR REPLACE TYPE ctx_feedback_type AS TABLE OF ctx_feedback_item_type;

CREATE OR REPLACE TYPE ctx_feedback_item_type AS OBJECT
(text        VARCHAR2(80),
  cardinality NUMBER,
  score       NUMBER,
  MAP MEMBER FUNCTION rank RETURN REAL,
  PRAGMA RESTRICT_REFERENCES (rank, RNDS, WNDS, RNPS, WNPS)
);

CREATE OR REPLACE TYPE BODY ctx_feedback_item_type AS
  MAP MEMBER FUNCTION rank RETURN REAL IS
  BEGIN
    RETURN score;
  END rank;
END;
END;
```

**See Also:** For an example of how to select from the HFEEDBACK table and its nested tables, refer to CTX_QUERY.HFEEDBACK in Chapter 11, "CTX_QUERY Package"

---

**CTX_DOC Result Tables**

The CTX_DOC procedures return results stored in a table. Before calling a procedure, you must create the table. The tables can be named anything, but must include columns with specific names and data types.

This section describes the following result tables and their required columns:

- Filter Table
- Gist Table
- Highlight Table
- Markup Table
- Theme Table

**Filter Table**

A filter table stores one row for each filtered document returned by CTX_DOC.FILTER. Filtered documents can be plain text or HTML.

When you call CTX_DOC.FILTER for a document, the document is processed through the filter defined for the text column and the results are stored in the filter table you specify.

Filter tables can be named anything, but must include the following columns, with names and datatypes as specified:
Gist Table

A Gist table stores one row for each Gist/theme summary generated by CTX_DOC.GIST.

Gist tables can be named anything, but must include the following columns, with names and data types as specified:

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUERY_ID</td>
<td>NUMBER</td>
<td>The identifier for the results generated by a particular call to CTX_DOC.FILTER (only populated when table is used to store results from multiple FILTER calls)</td>
</tr>
<tr>
<td>DOCUMENT</td>
<td>CLOB</td>
<td>Text of the document, stored in plain text or HTML.</td>
</tr>
</tbody>
</table>

Highlight Table

A highlight table stores offset and length information for highlighted terms in a document. This information is generated by CTX_DOC.HIGHLIGHT. Highlighted terms can be the words or phrases that satisfy a word or an ABOUT query.

If a document is formatted, the text is filtered into either plain text or HTML and the offset information is generated for the filtered text. The offset information can be used to highlight query terms for the same document filtered with CTX_DOC.FILTER.

Highlight tables can be named anything, but must include the following columns, with names and datatypes as specified:

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUERY_ID</td>
<td>NUMBER</td>
<td>The identifier for the results generated by a particular call to CTX_DOC.HIGHLIGHT (only populated when table is used to store results from multiple HIGHLIGHT calls)</td>
</tr>
<tr>
<td>OFFSET</td>
<td>NUMBER</td>
<td>The position of the highlight in the document, relative to the start of document which has a position of 1.</td>
</tr>
<tr>
<td>LENGTH</td>
<td>NUMBER</td>
<td>The length of the highlight.</td>
</tr>
</tbody>
</table>

Markup Table

A markup table stores documents in plain text or HTML format with the query terms in the documents highlighted by markup tags. This information is generated when you call CTX_DOC.MARKUP.
Markup tables can be named anything, but must include the following columns, with names and datatypes as specified:

**Table A–11 Markup Table**

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUERY_ID</td>
<td>NUMBER</td>
<td>The identifier for the results generated by a particular call to CTX_DOC.MARKUP (only populated when table is used to store results from multiple MARKUP calls)</td>
</tr>
<tr>
<td>DOCUMENT</td>
<td>CLOB</td>
<td>Marked-up text of the document, stored in plain text or HTML format</td>
</tr>
</tbody>
</table>

**Theme Table**

A theme table stores one row for each theme generated by CTX_DOC.THEMES. The value stored in the THEME column is either a single theme phrase or a string of parent themes, separated by colons.

Theme tables can be named anything, but must include the following columns, with names and data types as specified:

**Table A–12 Theme Table**

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUERY_ID</td>
<td>NUMBER</td>
<td>Query ID</td>
</tr>
<tr>
<td>THEME</td>
<td>VARCHAR2 (2000)</td>
<td>Theme phrase or string of parent themes separated by colons (:)</td>
</tr>
<tr>
<td>WEIGHT</td>
<td>NUMBER</td>
<td>Weight of theme phrase relative to other theme phrases for the document.</td>
</tr>
</tbody>
</table>

**Token Table**

A token table stores the text tokens for a document as output by the CTX_DOC.TOKENS procedure. Token tables can be named anything, but must include the following columns, with names and data types as specified:

**Table A–13 Token Table**

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUERY_ID</td>
<td>NUMBER</td>
<td>The identifier for the results generated by a particular call to CTX_DOC.HIGHLIGHT (only populated when table is used to store results from multiple HIGHLIGHT calls)</td>
</tr>
<tr>
<td>TOKEN</td>
<td>VARCHAR2 (64)</td>
<td>The token string in the text.</td>
</tr>
<tr>
<td>OFFSET</td>
<td>NUMBER</td>
<td>The position of the token in the document, relative to the start of document which has a position of 1.</td>
</tr>
<tr>
<td>LENGTH</td>
<td>NUMBER</td>
<td>The character length of the token.</td>
</tr>
</tbody>
</table>

**CTX_THES Result Tables and Data Types**

The CTX_THES expansion functions such as BT, NT, and SYN can return the expansions in a table of type EXP_TAB. Specify the name of your table with the restab argument.
EXP_TAB Table Type

The EXP_TAB table type is a table of rows of type EXP_REC.

The EXP_REC and EXP_TAB types are defined as follows in the CTXSYS schema:

type exp_rec is record {
    xrel varchar2(12),
    xlevel number,
    xphrase varchar2(256)
};

type exp_tab is table of exp_rec index by binary_integer;

When you call a thesaurus expansion function and specify RESTAB, the system returns the expansion as an EXP_TAB table. Each row in this table is of type EXP_REC and represents a word or phrase in the expansion. Table A–14 describes the fields in EXP_REC:

Table A–14  EXP_TAB Table Type (EXP_REC)

<table>
<thead>
<tr>
<th>EXP_REC Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>xrel</td>
<td>The xrel field contains the relation of the term to the input term (for example, 'SYN', 'PT', 'RT', and so on). The xrel value is PHRASE when the input term appears in the expansion. For translations, the xrel value is the language.</td>
</tr>
<tr>
<td>xlevel</td>
<td>The xlevel field is the level of the relation. This is used mainly when xrel is a hierarchical relation (BT*/NT*). The xlevel field is 0 when xrel is PHRASE. The xlevel field is 2 for translations of synonyms under TRSYN. The xlevel field is 1 for operators that are not hierarchical, such as PT and RT.</td>
</tr>
<tr>
<td>xphrase</td>
<td>The xphrase is the related term. This includes a qualifier in parentheses, if one exists for the related term. Compound terms are not de-compounded.</td>
</tr>
</tbody>
</table>
Oracle Text Supported Document Formats

This appendix contains a list of the document formats supported by the automatic (AUTO_FILTER) filtering technology. The following topics are covered in this appendix:

- About Document Filtering Technology
- Supported Document Formats

About Document Filtering Technology

The automatic filtering technology in Oracle Text uses Oracle Outside in HTML Export technology. This technology also enables you to convert documents to HTML for document presentation with the CTX_DOC package.

To use automatic filtering for indexing and DML processing, you must specify the AUTO_FILTER object in your filter preference.

To use automatic filtering technology for converting documents to HTML with the CTX_DOC package, you need not use the AUTO_FILTER indexing preference, but you must still set up your environment to use this filtering technology, as described in this appendix.

Note: The underlying technology used by Oracle Text was migrated to Oracle Outside in HTML Export in release 11.1.0.7. See "Formats No Longer Supported in 11.1.0.7" on page B-11 for a list of formats that are no longer supported as a result of this migration. Applications that require support for those formats can use USER_FILTER to plug in third-party filtering technology supporting those formats. See "USER_FILTER" on page 2-24 for more information.

Latest Updates for Patch Releases

The supported platforms and formats listed in this appendix apply for this release. These supported formats are updated for patch releases.

Restrictions on Format Support

Password-protected documents and documents with password-protected content are not supported by the AUTO_FILTER filter.

For other limitations, refer to sections in this chapter concerning specific document types.
Supported Platforms for AUTO_FILTER Document Filtering Technology

Several platforms can take advantage of AUTO_FILTER filter technology.

Supported Platforms

AUTO_FILTER filter technology is supported on the following platforms:

- Windows (x86 64-bit) Windows 2003 x64 Standard, Enterprise, and Datacenter Editions (64-bit Extended Systems)
- HP-UX (PA-RISC 64-bit) 11.i
- HP/UX (Itanium 64) 11i
- IBM AIX (32-bit pSeries) 5.1 - 5.3
- iSeries (OS/400 using PASE) V5R2
- Red Hat Linux (x86) Advanced Server 3, 4, and 5
- Red Hat Linux (x86) Red Hat Enterprise Linux (RHEL) 4
- Red Hat Linux (Itanium 64) Advanced Server 3, 4, and 5
- Red Hat Linux (zSeries, 31-bit) Advanced Server 3 and 4
- Red Hat Enterprise Linux AS/ES 3.0, 4.0 and 5.0, x86-64 (AMD64/EM64T)
- Oracle Enterprise Linux 4.0 and 5.0, x86-64 (AMD64/EM64T)
- SuSE Linux (x86) 9, 10, and Enterprise Server 9.0
- SuSE Linux (x86 64-bit) SUSE Enterprise Server (SLES) 9, 10
- SuSE Linux (Itanium 64) Enterprise Server 8
- SuSE Linux (zSeries, 31-bit) 9
- Sun Solaris (SPARC 64-bit) 9.x - 10.x
- Sun Solaris (x86-64-bit) 10x

Note that some of these platforms may not be supported by the Oracle Database.

Filtering on PDF Documents and Security Settings

A PDF document can have different levels of security settings as follows:

<table>
<thead>
<tr>
<th>Security Level</th>
<th>Description</th>
<th>PDF Version</th>
<th>Encryption</th>
<th>AUTO_FILTER Support Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Requires a password for opening the document.</td>
<td>1.2+</td>
<td>40 bit RC4</td>
<td>Not supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.4+</td>
<td>128 bit RC4</td>
<td>Not supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5+</td>
<td>128 bit RC4</td>
<td>Not supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.6+</td>
<td>128 bit AES</td>
<td>Not supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.7+</td>
<td>256 bit AES</td>
<td>Not supported.</td>
</tr>
</tbody>
</table>
PDF Filtering Limitations

The following limitations apply when filtering PDF files:

- Multi-byte PDFs are supported, provided the PDF document is created using Character ID-keyed (CID) fonts, predefined CJK CMap files, or ToUnicode font encodings, and the document does not contain embedded fonts.

- Embedded fonts in a PDF document are not filtered correctly. They are usually displayed using the question mark (?) replacement character.

- Hyperlinks in a PDF are not active when displayed in a browser or a viewing window.

- Annotations, such as notes, sound, or movies, are not supported.

Environment Variables

No environment variables need to be set by the user.

General Limitations

AUTO_FILTER filter technology has the following limitations:

- Any ASCII characters less than 0x20 (decimal 32) are converted to hexadecimal numbers.

- Files larger than 2GB are not handled.

Table B–1 (Cont.) AUTO_FILTER Behavior with PDF Security Settings

<table>
<thead>
<tr>
<th>Security Level</th>
<th>Description</th>
<th>PDF Version</th>
<th>Encryption</th>
<th>AUTO_FILTER Support Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 2</td>
<td>Disallows user printing of the document.</td>
<td>1.2+</td>
<td>40 bit RC4</td>
<td>Supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.4+</td>
<td>128 bit RC4</td>
<td>Supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5+</td>
<td>128 bit RC4</td>
<td>Supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.6+</td>
<td>128 bit AES</td>
<td>Not supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.7+</td>
<td>256 bit AES</td>
<td>Not supported.</td>
</tr>
<tr>
<td>Level 3</td>
<td>Disallows user modification or change of the document.</td>
<td>1.2+</td>
<td>40 bit RC4</td>
<td>Supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.4+</td>
<td>128 bit RC4</td>
<td>Supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5+</td>
<td>128 bit RC4</td>
<td>Supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.6+</td>
<td>128 bit RC4</td>
<td>Not supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.7+</td>
<td>256 bit AES</td>
<td>Not supported.</td>
</tr>
<tr>
<td>Level 4</td>
<td>Disallows the user from copying or extracting content from the document.</td>
<td>1.2+</td>
<td>40 bit RC4</td>
<td>Supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.4+</td>
<td>128 bit RC4</td>
<td>Supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5+</td>
<td>128 bit RC4</td>
<td>Supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.6+</td>
<td>128 bit AES</td>
<td>Not supported.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.7+</td>
<td>256 bit AES</td>
<td>Not supported.</td>
</tr>
</tbody>
</table>
Supported Document Formats

The tables in this section list the document formats that Oracle Text supports for filtering.

Document filtering is used for indexing, DML, and for converting documents to HTML with the CTX_DOC package.

---

**Note:** These lists do not represent the complete list of formats that Oracle Text is able to process. The USER_FILTER and PROCEDURE_FILTER enable Oracle Text to process any document format, provided an external filter exists that can filter to some textual format like plain-text, HTML, XML, and so forth.

---

Word Processing and Desktop Publishing Formats

<table>
<thead>
<tr>
<th>Format</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adobe FrameMaker (MIF)</td>
<td>Versions 3.0, 4.0, 5.0, and 6.0 and Japanese 3.0, 4.0, 5.0, and 6.0 (text only)</td>
</tr>
<tr>
<td>ANSI Text</td>
<td>7 and 8 bit</td>
</tr>
<tr>
<td>ASCII Text</td>
<td>7 and 8 bit</td>
</tr>
<tr>
<td>DEC WPS Plus (DX)</td>
<td>Versions through 3.1</td>
</tr>
<tr>
<td>DEC WPS Plus (WPL)</td>
<td>Versions through 4.1</td>
</tr>
<tr>
<td>DisplayWrite 2 and 3 (TXT)</td>
<td>All versions</td>
</tr>
<tr>
<td>EBCDIC</td>
<td>All versions</td>
</tr>
<tr>
<td>Enable</td>
<td>Versions 3.0, 4.0, and 4.5</td>
</tr>
<tr>
<td>First Choice</td>
<td>Versions through 3.0</td>
</tr>
<tr>
<td>Framework</td>
<td>Version 3.0</td>
</tr>
<tr>
<td>Hangul</td>
<td>Versions 97, 2002, and 2005</td>
</tr>
<tr>
<td>IBM FFT</td>
<td>All versions</td>
</tr>
<tr>
<td>IBM Revisable Form Text</td>
<td>All versions</td>
</tr>
<tr>
<td>IBM Writing Assistant</td>
<td>Version 1.01</td>
</tr>
<tr>
<td>Just System Ichitaro</td>
<td>Versions 4.x through 6.x, 8.x through 13.x and 2004</td>
</tr>
<tr>
<td>JustWrite</td>
<td>Versions through 3.0</td>
</tr>
<tr>
<td>Legacy</td>
<td>Versions 1.1</td>
</tr>
<tr>
<td>Lotus AMI/AMI Professional</td>
<td>Versions 3.1</td>
</tr>
<tr>
<td>Lotus Manuscript</td>
<td>Version 2.0</td>
</tr>
<tr>
<td>Lotus Word Pro (non-Windows)</td>
<td>Versions SmartSuite 97, Millennium, and Millennium 9.6 (text only)</td>
</tr>
<tr>
<td>Lotus Word Pro (Windows)</td>
<td>Versions SmartSuite 96, 97, and Millennium and Millennium 9.6</td>
</tr>
<tr>
<td>MacWrite II</td>
<td>Version 1.1</td>
</tr>
<tr>
<td>MASS11</td>
<td>Versions through 8.0</td>
</tr>
<tr>
<td>Microsoft Rich Text Format (RTF)</td>
<td>All versions</td>
</tr>
<tr>
<td>Format</td>
<td>Version</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Microsoft Word (DOS)</td>
<td>Versions through 6.0</td>
</tr>
<tr>
<td>Microsoft Word (Mac)</td>
<td>Versions 4.0 - 2004</td>
</tr>
<tr>
<td>Microsoft Word (Windows)</td>
<td>Versions through 2007</td>
</tr>
<tr>
<td>Microsoft WordPad</td>
<td>All versions</td>
</tr>
<tr>
<td>Microsoft Works (DOS)</td>
<td>Versions through 2.0</td>
</tr>
<tr>
<td>Microsoft Works (Mac)</td>
<td>Versions through 2.0</td>
</tr>
<tr>
<td>Microsoft Works (Windows)</td>
<td>Versions through 4.0</td>
</tr>
<tr>
<td>Microsoft Windows Write</td>
<td>Versions through 3.0</td>
</tr>
<tr>
<td>MultiMate</td>
<td>Versions through 4.0</td>
</tr>
<tr>
<td>Navy DIF</td>
<td>All versions</td>
</tr>
<tr>
<td>Nota Bene</td>
<td>Version 3.0</td>
</tr>
<tr>
<td>Novell Perfect Works</td>
<td>Version 2.0</td>
</tr>
<tr>
<td>Novell/Corel WordPerfect (DOS)</td>
<td>Versions through 6.1</td>
</tr>
<tr>
<td>Novell/Corel WordPerfect (Mac)</td>
<td>Versions 1.02 through 3.0</td>
</tr>
<tr>
<td>Novell/Corel WordPerfect (Windows)</td>
<td>Versions through 12.0</td>
</tr>
<tr>
<td>Office Writer</td>
<td>Versions 4.0 - 6.0</td>
</tr>
<tr>
<td>OpenOffice Writer (Windows and UNIX)</td>
<td>OpenOffice version 1.1 and 2.0</td>
</tr>
<tr>
<td>PC-File Letter</td>
<td>Versions through 5.0</td>
</tr>
<tr>
<td>PC-File+ Letter</td>
<td>Versions through 3.0</td>
</tr>
<tr>
<td>PFS:Write</td>
<td>Versions A, B, and C</td>
</tr>
<tr>
<td>Professional Write Plus (Windows)</td>
<td>Version 1.0</td>
</tr>
<tr>
<td>Q&amp;A (DOS)</td>
<td>Version 2.0</td>
</tr>
<tr>
<td>Q&amp;A Write (Windows)</td>
<td>Version 3.0</td>
</tr>
<tr>
<td>Samna Word</td>
<td>Versions through Samna Word IV+</td>
</tr>
<tr>
<td>Signature</td>
<td>Version 1.0</td>
</tr>
<tr>
<td>SmartWare II</td>
<td>Version 1.02</td>
</tr>
<tr>
<td>Sprint</td>
<td>Versions through 1.0</td>
</tr>
<tr>
<td>StarOffice Writer</td>
<td>Version 5.2 (text only) and 6.x through 8.x</td>
</tr>
<tr>
<td>Total Word</td>
<td>Version 1.2</td>
</tr>
<tr>
<td>Unicode Text</td>
<td>All versions</td>
</tr>
<tr>
<td>UTF-8</td>
<td>All versions</td>
</tr>
<tr>
<td>Volkswriter 3 and 4</td>
<td>Versions through 1.0</td>
</tr>
<tr>
<td>Wang PC (IWP)</td>
<td>Versions through 2.6</td>
</tr>
<tr>
<td>WordMARC</td>
<td>Versions through Composer Plus</td>
</tr>
<tr>
<td>WordStar (Windows)</td>
<td>Version 1.0</td>
</tr>
<tr>
<td>WordStar 2000 (DOS)</td>
<td>Versions through 3.0</td>
</tr>
<tr>
<td>XyWrite</td>
<td>Versions through III Plus</td>
</tr>
</tbody>
</table>
### Spreadsheet Formats

<table>
<thead>
<tr>
<th>Format</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable</td>
<td>Versions 3.0, 4.0, and 4.5</td>
</tr>
<tr>
<td>First Choice</td>
<td>Versions through 3.0</td>
</tr>
<tr>
<td>Framework</td>
<td>Version 3.0</td>
</tr>
<tr>
<td>Lotus 1-2-3 (DOS &amp; Windows)</td>
<td>Versions through 5.0</td>
</tr>
<tr>
<td>Lotus 1-2-3 (OS/2)</td>
<td>Versions through 2.0</td>
</tr>
<tr>
<td>Lotus 1-2-3 Charts (DOS &amp; Windows)</td>
<td>Versions through 5.0</td>
</tr>
<tr>
<td>Lotus 1-2-3 for SmartSuite</td>
<td>Versions 97 - Millennium 9.6</td>
</tr>
<tr>
<td>Lotus Symphony</td>
<td>Versions 1.0, 1.1, and 2.0</td>
</tr>
<tr>
<td>Mac Works</td>
<td>Version 2.0</td>
</tr>
<tr>
<td>Microsoft Excel Charts</td>
<td>Versions 2.x - 7.0</td>
</tr>
<tr>
<td>Microsoft Excel (Windows)</td>
<td>Versions 2.2 through 2007</td>
</tr>
<tr>
<td>Microsoft Multiplan</td>
<td>Version 4.0</td>
</tr>
<tr>
<td>Microsoft Works (Windows)</td>
<td>Versions through 4.0</td>
</tr>
<tr>
<td>Microsoft Works (DOS)</td>
<td>Versions through 2.0</td>
</tr>
<tr>
<td>Microsoft Works (Mac)</td>
<td>Versions through 2.0</td>
</tr>
<tr>
<td>Mosaic Twin</td>
<td>Version 2.5</td>
</tr>
<tr>
<td>Novell Perfect Works</td>
<td>Version 2.0</td>
</tr>
<tr>
<td>PFS:Professional Plan</td>
<td>Version 1.0</td>
</tr>
<tr>
<td>Quattro Pro (DOS)</td>
<td>Versions through 5.0 (text only)</td>
</tr>
<tr>
<td>Quattro Pro (Windows)</td>
<td>Version through 12.0 (text only)</td>
</tr>
<tr>
<td>SmartWare II</td>
<td>Version 1.02</td>
</tr>
<tr>
<td>StarOffice/OpenOffice Calc (Windows and UNIX)</td>
<td>StarOffice versions 5.2 (text only) through 8.x and OpenOffice version 1.1 and 2.0</td>
</tr>
<tr>
<td>SuperCalc 5</td>
<td>Version 4.0</td>
</tr>
<tr>
<td>VP Planner 3D</td>
<td>Version 1.0</td>
</tr>
</tbody>
</table>

### Presentation Formats

<table>
<thead>
<tr>
<th>Format</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corel / Novell Presentations</td>
<td>Versions through 12.0</td>
</tr>
<tr>
<td>Harvard Graphics (DOS)</td>
<td>Versions 2.x and 3.x</td>
</tr>
<tr>
<td>Harvard Graphics (Windows)</td>
<td>Windows versions</td>
</tr>
<tr>
<td>Freelance (Windows)</td>
<td>Versions through Millennium 9.6</td>
</tr>
<tr>
<td>Freelance (OS/2)</td>
<td>Versions through 2.0</td>
</tr>
<tr>
<td>Microsoft PowerPoint (Windows)</td>
<td>Versions 3.0 through 2007</td>
</tr>
<tr>
<td>Microsoft PowerPoint (Mac)</td>
<td>Versions 4.0 through v.x</td>
</tr>
</tbody>
</table>
Database Formats

<table>
<thead>
<tr>
<th>Format</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>Versions through 2.0</td>
</tr>
<tr>
<td>dBASE</td>
<td>Versions through 5.0</td>
</tr>
<tr>
<td>DataEase</td>
<td>Version 4.x</td>
</tr>
<tr>
<td>dBXL</td>
<td>Version 1.3</td>
</tr>
<tr>
<td>Enable</td>
<td>Versions 3.0, 4.0, and 4.5</td>
</tr>
<tr>
<td>First Choice</td>
<td>Versions through 3.0</td>
</tr>
<tr>
<td>FoxBase</td>
<td>Version 2.1</td>
</tr>
<tr>
<td>Framework</td>
<td>Version 3.0</td>
</tr>
<tr>
<td>Microsoft Works (Windows)</td>
<td>Versions through 4.0</td>
</tr>
<tr>
<td>Microsoft Works (DOS)</td>
<td>Versions through 2.0</td>
</tr>
<tr>
<td>Microsoft Works (Mac)</td>
<td>Versions through 2.0</td>
</tr>
<tr>
<td>Paradox (DOS)</td>
<td>Versions through 4.0</td>
</tr>
<tr>
<td>Paradox (Windows)</td>
<td>Versions through 1.0</td>
</tr>
<tr>
<td>Personal R:BASE</td>
<td>Version 1.0</td>
</tr>
<tr>
<td>R:BASE 5000</td>
<td>Versions through 3.1</td>
</tr>
<tr>
<td>R:BASE System V</td>
<td>Version 1.0</td>
</tr>
<tr>
<td>Reflex</td>
<td>Version 2.0</td>
</tr>
<tr>
<td>Q &amp; A</td>
<td>Versions through 2.0</td>
</tr>
<tr>
<td>SmartWare II</td>
<td>Version 1.02</td>
</tr>
</tbody>
</table>

Archive File Format

When filtering an archive file, all the contents of the files inside the archive will be exported to a single output file. This will also include the contents of all subfolders and files inside the archive file.

Table B–2 lists the archive formats that Oracle supports.

<table>
<thead>
<tr>
<th>Format</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>GZIP</td>
<td></td>
</tr>
<tr>
<td>Microsoft Binder</td>
<td>Versions 7.0 - 97 (conversion of files contained in the Binder File is supported only on Windows)</td>
</tr>
<tr>
<td>UUEncode</td>
<td></td>
</tr>
<tr>
<td>UNIX Compress</td>
<td></td>
</tr>
<tr>
<td>UNIX Tar</td>
<td></td>
</tr>
</tbody>
</table>
Email Formats

- MIME formats
  - EML
  - MHT (Web Archive)
  - NWS (Newsgroup single-part and multi-part)
  - Simple Text Mail (defined in RFC 2822)
- TNEF format
- MIME encodings, including
  - base64 (defined in RFC 1521)
  - binary (defined in RFC 1521)
  - binhex (defined in RFC 1741)
  - btoa
  - quoted-printable (defined in RFC 1521)
  - utf-7 (defined in RFC 2152)
  - uue
  - xxe
  - yenc

In addition, the body of a message can be encoded in several ways. The following encodings are supported:
- HTML
- RTF
- TNEF

MIME Support Notes
The following formats are supported:

Table B–2 (Cont.) Supported Archive File Formats

<table>
<thead>
<tr>
<th>Format</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZIP</td>
<td>PKWARE versions through 2.04g</td>
</tr>
<tr>
<td>LZA Self-Extracting</td>
<td>Compress</td>
</tr>
<tr>
<td>LZH Compress</td>
<td></td>
</tr>
</tbody>
</table>

Table B–2 (Cont.) Supported Archive File Formats

<table>
<thead>
<tr>
<th>Format</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIME</td>
<td>MIME-encoded mail messages.</td>
</tr>
</tbody>
</table>
- Text/enriched (defined in RFC 1523)
- Text/richtext (defined in RFC1341)
- Embedded mail message (defined in RFC 822) - this is handled as a link to a new message

The attachments of a MIME message can be stored in many formats. Oracle Corporation processes all attachment types that its technology supports.

### Other Formats

<table>
<thead>
<tr>
<th>Format</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executable (EXE, DLL)</td>
<td>Versions through 3.0, with some limitations</td>
</tr>
<tr>
<td>HTML</td>
<td>Macromedia Flash 6.x, Macromedia Flash 7.x, and Macromedia Flash Lite (text only)</td>
</tr>
<tr>
<td>MacroMedia Flash</td>
<td></td>
</tr>
<tr>
<td>Microsoft Project</td>
<td>Versions 98 - 2003 (text only)</td>
</tr>
<tr>
<td>MP3</td>
<td>ID3 information</td>
</tr>
<tr>
<td>vCard, vCalendar</td>
<td>Version 2.1</td>
</tr>
<tr>
<td>Windows Executable</td>
<td></td>
</tr>
<tr>
<td>WML</td>
<td>Version 5.2</td>
</tr>
<tr>
<td>XML</td>
<td>Text only</td>
</tr>
<tr>
<td>Yahoo Instant</td>
<td></td>
</tr>
</tbody>
</table>

### Graphic Formats

Table B–3 lists the graphic formats that the AUTO_FILTER filter recognizes. This means that indexing a text column that contains any of these formats produces no error. As such, it is safe for the column to contain any of these formats.

Formats are categorized as either *embedded graphics* or *standalone graphics*. Embedded graphics are inserted or referenced within a document.

---

**Note:** The AUTO_FILTER filter cannot extract textual information from graphics.

---

### Table B–3  Supported Graphics Formats for AUTO_FILTER Filter

<table>
<thead>
<tr>
<th>Graphics Format</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adobe Photoshop (PSD)</td>
<td>Version 4.0</td>
</tr>
<tr>
<td>Windows Icon Cursor (ICO)</td>
<td>no specific version</td>
</tr>
<tr>
<td>Adobe Photoshop (PSD)</td>
<td>Version 4.0</td>
</tr>
<tr>
<td>Adobe Illustrator</td>
<td>Versions 7.0 and 9.0</td>
</tr>
<tr>
<td>Adobe FrameMaker graphics (FMV)</td>
<td>Vector/raster through 5.0</td>
</tr>
<tr>
<td>Adobe Acrobat (PDF)</td>
<td>Versions 1.0, 2.1, 3.0, 4.0, 5.0, 6.0, and 7.0 (including Japanese PDF)</td>
</tr>
<tr>
<td>Ami Draw (SDW)</td>
<td>Ami Draw</td>
</tr>
<tr>
<td>Graphics Format</td>
<td>Version</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>AutoCAD Interchange and Native Drawing formats (DXF and DWG)</td>
<td>AutoCAD Drawing Versions 2.5 - 2.6, 9.0-14.0, 2000i and 2002</td>
</tr>
<tr>
<td>AutoShadeRendering (RND)</td>
<td>Version 2.0</td>
</tr>
<tr>
<td>Binary Group 3 Fax</td>
<td>All versions</td>
</tr>
<tr>
<td>Bitmap (BMP, RLE, ICO, CUR, OS/2 DIB, and WARP)</td>
<td>All versions</td>
</tr>
<tr>
<td>CALS Raster (GP4)</td>
<td>Type I and Type II</td>
</tr>
<tr>
<td>Corel Clipart format (CMX)</td>
<td>Versions 5 through 6</td>
</tr>
<tr>
<td>Corel Draw (CDR)</td>
<td>Versions 3.x - 8.x</td>
</tr>
<tr>
<td>Corel Draw (CDR with TIFF header)</td>
<td>Versions 2.x - 9.x</td>
</tr>
<tr>
<td>Computer Graphics Metafile (CGM)</td>
<td>ANSI, CALS NIST version 3.0</td>
</tr>
<tr>
<td>Encapsulated PostScript (EPS)</td>
<td>TIFF header only</td>
</tr>
<tr>
<td>GEM Paint (IMG)</td>
<td>All versions</td>
</tr>
<tr>
<td>Graphics Environment Mgr (GEM)</td>
<td>Bitmap and vector</td>
</tr>
<tr>
<td>Graphics Interchange Format (GIF)</td>
<td>All versions</td>
</tr>
<tr>
<td>Hewlett Packard Graphics Language (HPGL)</td>
<td>Version 2.0</td>
</tr>
<tr>
<td>IBM Graphics Data Format (GDF)</td>
<td>Version 1.0</td>
</tr>
<tr>
<td>IBM Picture Interchange Format (PIF)</td>
<td>Version 1.0</td>
</tr>
<tr>
<td>Initial Graphics Exchange Spec (IGES)</td>
<td>Version 5.1</td>
</tr>
<tr>
<td>JBIG2</td>
<td>JBIG2 graphic embeddings in PDF files</td>
</tr>
<tr>
<td>JFIF (JPEG not in TIFF format)</td>
<td>All versions</td>
</tr>
<tr>
<td>JPEG (including EXIF)</td>
<td>All versions</td>
</tr>
<tr>
<td>Kodak Flash Pix (FPX)</td>
<td>All versions</td>
</tr>
<tr>
<td>Kodak Photo CD (PCD)</td>
<td>Version 1.0</td>
</tr>
<tr>
<td>Lotus PIC</td>
<td>All versions</td>
</tr>
<tr>
<td>Lotus Snapshot</td>
<td>All versions</td>
</tr>
<tr>
<td>Macintosh PIC1 and PICT2</td>
<td>Bitmap only</td>
</tr>
<tr>
<td>MacPaint (PNTG)</td>
<td>All versions</td>
</tr>
<tr>
<td>Micrografx Draw (DRW)</td>
<td>Versions through 4.0</td>
</tr>
<tr>
<td>Micrografx Designer (DRW)</td>
<td>Versions through 3.1</td>
</tr>
<tr>
<td>Micrografx Designer (DFS)</td>
<td>Windows 95, version 6.0</td>
</tr>
<tr>
<td>Novell PerfectWorks (Draw)</td>
<td>Version 2.0</td>
</tr>
<tr>
<td>OS/2 PM Metafile (MET)</td>
<td>Version 3.0</td>
</tr>
<tr>
<td>Paint Shop Pro 6 (PSP)</td>
<td>Windows only, versions 5.0 - 6.0</td>
</tr>
<tr>
<td>PC Paintbrush (PCX and DCX)</td>
<td>All versions</td>
</tr>
<tr>
<td>Portable Bitmap (PBM)</td>
<td>All versions</td>
</tr>
<tr>
<td>Portable Graymap (PGM)</td>
<td>No specific version</td>
</tr>
</tbody>
</table>
**Graphics Formats Limitations**

AutoCAD drawing files are not supported on IBM AIX.

**Formats No Longer Supported in 11.1.0.7**

Certain document formats are not supported if you upgrade from release 11.1.0.6 to 11.1.0.7. This is because Oracle Text filtering technology has been migrated to Oracle Outside in HTML Export technology. To filter these unsupported formats, you can plug in a third party filtering technology using `USER_FILTER`. See "USER_FILTER" on page 2-24 for more information.

Table B–4 lists the formats supported in release 11.1.0.6, but not in 11.1.0.7.

**Table B–4 Formats Supported in Release 11.1.0.6 and not in 11.1.0.7**

<table>
<thead>
<tr>
<th>Format</th>
<th>Versions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Processing Formats</td>
<td></td>
</tr>
<tr>
<td>Applix Words (AW)</td>
<td>3.11, 4.0, 4.1, 4.2, 4.3, 4.4</td>
</tr>
<tr>
<td>JustSystems Ichitaro (JTD)</td>
<td>2005</td>
</tr>
<tr>
<td>Folio Flat File (FFF)</td>
<td>3.1</td>
</tr>
<tr>
<td>Fujitsu Oasys (OA2)</td>
<td>7</td>
</tr>
<tr>
<td>Lotus Word Pro (LWP)</td>
<td>9.7, 9.8</td>
</tr>
<tr>
<td>WordPerfect for Linux</td>
<td>All versions</td>
</tr>
</tbody>
</table>
**Table B-4 (Cont.) Formats Supported in Release 11.1.0.6 and not in 11.1.0.7**

<table>
<thead>
<tr>
<th>Format</th>
<th>Versions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desktop Publishing Formats</td>
<td></td>
</tr>
<tr>
<td>Adobe Framemaker (MIF)</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Spreadsheet Formats</td>
<td></td>
</tr>
<tr>
<td>Applix Spreadsheets (AS)</td>
<td>4.2, 4.3, 4.4</td>
</tr>
<tr>
<td>Lotus 1-2-3 (123)</td>
<td>Millennium Edition R9, 9.8</td>
</tr>
<tr>
<td>Microsoft Works Spreadsheet (DOS)</td>
<td>3.4</td>
</tr>
<tr>
<td>Microsoft Works Spreadsheet (Mac)</td>
<td>3.4</td>
</tr>
<tr>
<td>Comma-Separated Values (SCV)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentation Formats</td>
<td></td>
</tr>
<tr>
<td>Applix Presents (AG)</td>
<td>4.0, 4.2, 4.3, 4.4</td>
</tr>
<tr>
<td>Lotus Freelance Graphics (PRE)</td>
<td>Millennium Edition R9, 9.8</td>
</tr>
<tr>
<td>Microsoft Visio XML Format</td>
<td>2003</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Graphic Formats</td>
<td></td>
</tr>
<tr>
<td>SGI RGB Image (RGB)</td>
<td>No specific version</td>
</tr>
<tr>
<td>Windows Animated Cursor (ANI)</td>
<td>No specific version</td>
</tr>
<tr>
<td>WordPerfect Graphics 2 (WPG2)</td>
<td>7</td>
</tr>
<tr>
<td>Microsoft Office Drawing (MSO)</td>
<td>No specific version</td>
</tr>
<tr>
<td>Windows Icon Cursor (ICO)</td>
<td>No specific version</td>
</tr>
</tbody>
</table>
This appendix provides examples of how to load text into a text column, and the structure of ctxload import files. This appendix contains these topics:

- SQL INSERT Example
- SQL*Loader Example
- Structure of ctxload Thesaurus Import File

### SQL INSERT Example

A simple way to populate a text table is to create a table with two columns, `id` and `text`, using `CREATE TABLE` and then use the `INSERT` statement to load the data. This example makes the `id` column the primary key, which is optional. The `text` column is `VARCHAR2`:

```sql
CREATE TABLE docs (id NUMBER PRIMARY KEY, text VARCHAR2(80));
```

To populate the `text` column, use the `INSERT` statement as follows:

```sql
INSERT INTO docs VALUES(1, 'this is the text of the first document');
INSERT INTO docs VALUES(12, 'this is the text of the second document');
```

### SQL*Loader Example

The following example shows how to use SQL*Loader to load mixed format documents from the operating system to a `BLOB` column. The example has two steps:

- Create the table
- Enter the SQL*Loader command that reads control file and loads data into table

**See Also:** For a complete discussion on using SQL*Loader, see Oracle Database Utilities

### Creating the Table

This example loads a table `articles_formatted` created as follows:

```sql
CREATE TABLE articles_formatted (
  ARTICLE_ID   NUMBER PRIMARY KEY ,
  AUTHOR       VARCHAR2(30),
  FORMAT       VARCHAR2(30),
  PUB_DATE     DATE,
  TITLE        VARCHAR2(256),
  TEXT         BLOB
);```
The `article_id` column is the primary key. Documents are loaded in the `text` column, which is of type `BLOB`.

### Issuing the SQL*Loader Command

The following command starts the loader, which reads the control file `LOADER1.DAT`:

```
sqlldr userid=demo/demo control=loader1.dat log=loader.log
```

**Example Control File: `loader1.dat`**

This SQL*Loader control file defines the columns to be loaded and instructs the loader to load the data line by line from `loader2.dat` into the `articles_formatted` table. Each line in `loader2.dat` holds a comma separated list of fields to be loaded.

```sql
-- load file example
load data
INFILE 'loader2.dat'
INTO TABLE articles_formatted
APPEND
FIELDS TERMINATED BY ','
(article_id SEQUENCE (MAX,1),
 author CHAR(30),
 format,
 pub_date SYSDATE,
 title,
 ext_fname FILLER CHAR(80),
 text LOBFILE(ext_fname) TERMINATED BY EOF)
```

This control file instructs the loader to load data from `loader2.dat` to the `articles_formatted` table in the following way:

1. The ordinal position of the line describing the document fields in `loader2.dat` is written to the `article_id` column.
2. The first field on the line is written to the `author` column.
3. The second field on the line is written to the `format` column.
4. The current date given by `SYSDATE` is written to the `pub_date` column.
5. The title of the document, which is the third field on the line, is written to the `title` column.
6. The name of each document to be loaded is read into the `ext_fname` temporary variable, and the actual document is loaded in the `text` BLOB column.

**Example Data File: `loader2.dat`**

This file contains the data to be loaded into each row of the table, `articles_formatted`.

Each line contains a comma separated list of the fields to be loaded in `articles_formatted`. The last field of every line names the file to be loaded in to the text column:

```
Ben Kanobi, plaintext,Kawasaki news article,../sample_docs/kawasaki.txt,
Joe Bloggs, plaintext,Java plug-in,../sample_docs/javaplugin.txt,
John Hancock, plaintext,Declaration of Independence,../sample_docs/indep.txt,
M. S. Developer, Word7, Newsletter example,../sample_docs/newsletter.doc,
M. S. Developer, Word7, Resume example,../sample_docs/resume.doc,
```
X. L. Developer, Excel7, Common example, ../sample_docs/common.xls,
X. L. Developer, Excel7, Complex example, ../sample_docs/solvsamp.xls,
Paw R. Point, Powerpoint7, Generic presentation, ../sample_docs/generic.ppt,
Paw R. Point, Powerpoint7, Meeting presentation, ../sample_docs/meeting.ppt,
Java Man, PDF, Java Beans paper, ../sample_docs/j_bean.pdf,
Java Man, PDF, Java on the server paper, ../sample_docs/j_svr.pdf,
Ora Webmaster, HTML, Oracle home page, ../sample_docs/oramnu97.html,
Ora Webmaster, HTML, Oracle Company Overview, ../sample_docs/oraoverview.html,
John Constable, GIF, Laurence J. Ellison: Portrait, ../sample_docs/larry.gif,
Alan Greenspan, GIF, Oracle revenues: Graph, ../sample_docs/oragraph97.gif,
Giorgio Armani, GIF, Oracle Revenues: Trend, ../sample_docs/oratrend.gif,

Structure of ctxload Thesaurus Import File

The import file must use the following format for entries in the thesaurus:

- **phrase**
  - **BT** broader_term
  - **NT** narrower_term1
  - **NT** narrower_term2
  - ...
  - **NT** narrower_termN
  - **BTG** broader_term
  - **NTG** narrower_term1
  - **NTG** narrower_term2
  - ...
  - **NTG** narrower_termN
  - **BTP** broader_term
  - **NTP** narrower_term1
  - **NTP** narrower_term2
  - ...
  - **NTP** narrower_termN
  - **BTI** broader_term
  - **NTI** narrower_term1
  - **NTI** narrower_term2
  - ...
  - **NTI** narrower_termN
  - **SYN** synonym1
  - **SYN** synonym2
  - ...
  - **SYN** synonymN
  - **USE** synonym1 or **SEE** synonym1 or **PT** synonym1
  - **RT** related_term1
  - **RT** related_term2
  - ...
  - **RT** related_termN
  - **SN** text

- **language_key**: term

**phrase**

is a word or phrase that is defined as having synonyms, broader terms, narrower terms, or related terms.
In compliance with ISO-2788 standards, a TT marker can be placed before a phrase to indicate that the phrase is the top term in a hierarchy; however, the TT marker is not required. In fact, ctxload ignores TT markers during import.

A top term is identified as any phrase that does not have a broader term (BT, BTG, BTP, or BTI).

**Note:** The thesaurus query operators (SYN, PT, BT, BTG, BTP, BTI, NT, NTG, NTP, NTI, and RT) are reserved words and, thus, cannot be used as phrases in thesaurus entries.

**BT, BTG, BTP, BTI broader_termN**

are the markers that indicate broader_termN is a broader (generic | partitive | instance) term for phrase.

broader_termN is a word or phrase that conceptually provides a more general description or category for phrase. For example, the word elephant could have a broader term of land mammal.

**NT, NTG, NTP, NTI narrower_termN**

are the markers that indicate narrower_termN is a narrower (generic | partitive | instance) term for phrase.

If phrase does not have a broader (generic | partitive | instance) term, but has one or more narrower (generic | partitive | instance) terms, phrase is created as a top term in the respective hierarchy (in an Oracle Text thesaurus, the BT/NT, BTG/NTG, BTP/NTP, and BTI/NTI hierarchies are separate structures).

narrower_termN is a word or phrase that conceptually provides a more specific description for phrase. For example, the word elephant could have a narrower terms of indian elephant and african elephant.

**SYN synonymN**

is a marker that indicates phrase and synonymN are synonyms within a synonym ring.

synonymN is a word or phrase that has the same meaning for phrase. For example, the word dog could have a synonym of canine.

**Note:** Synonym rings are not defined explicitly in Oracle Text thesauri. They are created by the transitive nature of synonyms.

**USE SEE PT synonym1**

are markers that indicate phrase and synonym1 are synonyms within a synonym ring (similar to SYN).

The markers USE, SEE or PT also indicate synonym1 is the preferred term for the synonym ring. Any of these markers can be used to define the preferred term for a synonym ring.
**Note:** If the user-defined thesaurus is to be used for compiling the Knowledge Base, then you must specify the preferred term when a synonym ring is declared. Use one of the keywords USE, SEE, or PT to specify which synonym to use when reporting query matches. Only one term can be a preferred term.

Not using one of these keywords may result in the failure to return results defined by a word's synonym. When compiling two or more thesauri that declare elements of the same synonym ring, the preferred term must be the same in both files, which ensures that only one word is defined as the preferred word in a synonym ring.

---

**RT related_termN**

is the marker that indicates related_termN is a related term for phrase.

related_termN is a word or phrase that has a meaning related to, but not necessarily synonymous with phrase. For example, the word dog could have a related term of wolf.

---

**Note:** Related terms are not transitive. If a phrase has two or more related terms, the terms are related only to the parent phrase and not to each other.

---

**SN text**

is the marker that indicates the following text is a scope note (for example, comment) for the preceding entry.

**language_key term**

term is the translation of phrase into the language specified by language_key.

---

**Alternate Hierarchy Structure**

In compliance with thesauri standards, the load file supports formatting hierarchies (BT/NT, BTG/NTG, BTP, NTP, BTI/NTI) by indenting the terms under the top term and using NT (or NTG, NTP, NTI) markers that include the level for the term:

```
phrase
  NT1 narrower_term1
  NT2 narrower_term1.1
  NT2 narrower_term1.2
    NT3 narrower_term1.2.1
    NT3 narrower_term1.2.2
  NT1 narrower_term2
  ...
  NT1 narrower_termN
```

Using this method, the entire branch for a top term can be represented hierarchically in the load file.

---

**Usage Notes for Terms in Import Files**

The following conditions apply to the structure of the entries in the import file:

- Each entry (phrase, BT, NT, or SYN) must be on a single line followed by a newline character.
- Entries can consist of a single word or phrases.
The maximum length of an entry (phrase, BT, NT, or SYN) is 255 bytes, not including the BT, NT, and SYN markers or the newline characters.

Entries cannot contain parentheses or plus signs.

Each line of the file that starts with a relationship (BT, NT, and so on) must begin with at least one space.

A phrase can occur more than once in the file.

Each phrase can have one or more narrower term entries (NT, NTG, NTP), broader term entries (BT, BTG, BTP), synonym entries, and related term entries.

Each broader term, narrower term, synonym, and preferred term entry must start with the appropriate marker and the markers must be in capital letters.

The broader terms, narrower terms, and synonyms for a phrase can be in any order.

Homographs must be followed by parenthetical disambiguators everywhere they are used.

For example: cranes (birds), cranes (lifting equipment)

Compound terms are signified by a plus sign between each factor (for example, buildings + construction).

Compound terms are allowed only as synonyms or preferred terms for other terms, never as terms by themselves, or in hierarchical relations.

Terms can be followed by a scope note (SN), total maximum length of 2000 bytes, on subsequent lines.

Multi-line scope notes are allowed, but require an SN marker on each line of the note.

Example of Incorrect SN usage:

<table>
<thead>
<tr>
<th>VIEW CAMERAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN Cameras with through-the lens focusing and a range of movements of the lens plane relative to the film plane</td>
</tr>
</tbody>
</table>

Example of Correct SN usage:

<table>
<thead>
<tr>
<th>VIEW CAMERAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN Cameras with through-the lens focusing and a SN range of movements of the lens plane relative SN to the film plane</td>
</tr>
</tbody>
</table>

Multi-word terms cannot start with reserved words (for example, use is a reserved word, so use other door is not an allowed term; however, use is an allowed term).

Usage Notes for Relationships in Import Files

The following conditions apply to the relationships defined for the entries in the import file:

- related term entries must follow a phrase or another related term entry
- related term entries start with one or more spaces, the RT marker, followed by white space, then the related term on the same line
- multiple related terms require multiple RT markers

Example of incorrect RT usage:
MOVING PICTURE CAMERAS
RT CINE CAMERAS
TELEVISION CAMERAS

Example of correct RT usage:
MOVING PICTURE CAMERAS
RT CINE CAMERAS
RT TELEVISION CAMERAS

- Terms are allowed to have multiple broader terms, narrower terms, and related terms

Examples of Import Files
This section provides three examples of correctly formatted thesaurus import files.

Example 1 (Flat Structure)
cat
SYN feline
NT domestic cat
NT wild cat
BT mammal
mammal
BT animal
domestic cat
NT Persian cat
NT Siamese cat
wild cat
NT tiger
tiger
NT Bengal tiger
dog
BT mammal
NT domestic dog
NT wild dog
SYN canine
domestic dog
NT German Shepherd
wild dog
NT Dingo

Example 2 (Hierarchical)
animal
   NT1 mammal
   NT2 cat
      NT3 domestic cat
      NT4 Persian cat
      NT4 Siamese cat
      NT3 wild cat
      NT4 tiger
      NT5 Bengal tiger
   NT2 dog
      NT3 domestic dog
      NT4 German Shepherd
      NT3 wild dog
      NT4 Dingo
cat
SYN feline
dog
SYN canine

**Example 3**

35MM CAMERAS
  BT MINIATURE CAMERAS
CAMERAS
  BT OPTICAL EQUIPMENT
NT MOVING PICTURE CAMERAS
NT STEREO CAMERAS
LAND CAMERAS
USE VIEW CAMERAS
VIEW CAMERAS
SN Cameras with through-the-lens focusing and a range of
SN movements of the lens plane relative to the film plane
UF LAND CAMERAS
BT STILL CAMERAS
This Appendix describes the multilingual features of Oracle Text. The following topics are discussed:

- Introduction
- Indexing
- Querying
- Supplied Stop Lists
- Knowledge Base
- Multilingual Features Matrix

Introduction

This appendix summarizes the main multilingual features for Oracle Text.

For a complete list of Oracle Globalization Support languages and character set support, refer to the Oracle Database Globalization Support Guide.

Indexing

The following sections describe the multilingual indexing features:

- Multilingual Features for Text Index Types
- Lexer Types
- Auto Lexer Features
- Basic Lexer Features
- Multi Lexer Features
- World Lexer Features

Multilingual Features for Text Index Types

The following sections describes the supported multilingual features for the Oracle Text index types.

See Also: "Lexer Types" on page D-2 for a description of available lexers
**CONTEXT Index Type**

The CONTEXT index type fully supports multilingual features, including use of the language and character set columns. The following lexers are supported:

- AUTO_LEXER
- MULTI_LEXER
- USER_LEXER
- WORLD_LEXER

CONTEXT also supports use of all Chinese, Japanese, and Korean language lexers as follows:

- CHINESE_LEXER
- CHINESE_VGRAM_LEXER
- JAPANESE_LEXER
- JAPANESE_VGRAM_LEXER
- KOREAN_MORPH_LEXER

**CTXCAT Index Type**

CTXCAT supports the multilingual features of the BASIC_LEXER with the exception of indexing themes, and supports the following additional lexers:

- AUTO_LEXER
- USER_LEXER
- WORLD_LEXER

CTXCAT also supports the following lexers:

- CHINESE_LEXER
- CHINESE_VGRAM_LEXER
- JAPANESE_LEXER
- JAPANESE_VGRAM_LEXER
- KOREAN_MORPH_LEXER

**CTXRULE Index Type**

The CTXRULE index type supports the multilingual features of the BASIC_LEXER including ABOUT and STEM operators. It also supports Japanese, Chinese, and Korean (when used with the SVM_CLASSIFIER).

**Lexer Types**

Oracle Text supports the indexing of different languages by enabling you to choose a lexer in the indexing process. The lexer you employ determines the languages you can index. Table D–1 describes the supported lexers.

<table>
<thead>
<tr>
<th>Lexer</th>
<th>Supported Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTO_LEXER</td>
<td>Lexer for indexing columns that contain documents of different languages.</td>
</tr>
</tbody>
</table>
Auto Lexer Features

AUTO_LEXER automatically detects document language, and performs language identification, word segmentation, and stemming. The AUTO_LEXER also enables customization of these components.

See Also: "AUTO_LEXER" on page 2-31

Basic Lexer Features

The following features are supported with the BASIC_LEXER preference. Enable these features with attributes of the BASIC_LEXER. Features such as alternate spelling, composite, and base letter can be enabled together for better search results.

Theme Indexing

Enables the indexing and subsequent querying of document concepts with the ABOUT operator with CONTEXT index types. These concepts are derived from the Oracle Text knowledge base. This feature is supported for English and French.

This feature is not supported with CTXCAT index types.

Alternate Spelling

This feature enables you to search on alternate spellings of words. For example, with alternate spelling enabled in German, a query on gross returns documents that contain groß and gross.

Oracle Text Multilingual Features  D-3
**Indexing**

This feature is supported in German, Danish, and Swedish.

Additionally, German can be indexed according to both traditional and reformed spelling conventions.

**See Also:** "Alternate Spelling" on page 16-2 and "New German Spelling" on page 16-2.

**Base Letter Conversion**

This feature enables you to query words with or without diacritical marks such as tildes, accents, and umlauts. For example, with a Spanish base-letter index, a query of *energia* matches documents containing both *energía* and *energia*.

This feature is supported for English and all other supported whitespace delimited languages. In English and French, you can use the basic lexer to enable theme indexing.

**See Also:** "Base-Letter Conversion" on page 16-2

**Composite**

This feature enables you to search on words that contain the specified term as a sub-composite. You must use the stem ($) operator. This feature is supported for German and Dutch.

For example, in German, a query of `$register` finds documents that contain *Bruttoregistertonne* and *Registertonne*.

**Index stems**

This feature enables you to specify a stemmer for stem indexing. Tokens are stemmed to a single base form at index time in addition to the normal forms. Specifying index stems enables better query performance for stem queries, for example `$computed`.

This feature is supported for English, Dutch, French, German, Italian, and Spanish.

**Multi Lexer Features**

The **MULTI_LEXER** lexer enables you to index a column that contains documents of different languages. During indexing Oracle Text examines the language column and switches in the language-specific lexer to process the document. Define the lexer preferences for each language before indexing.

The multi lexer enables you to set different preferences for languages. For example, you can have `composite` set to `TRUE` for German documents and `composite` set to `FALSE` for Dutch documents.

**World Lexer Features**

Like **MULTI_LEXER**, the **WORLD_LEXER** lexer enables you to index documents that contain different languages. It automatically detects the languages of a document and, therefore, does not require you to create a language column in the base table.

**WORLD_LEXER** processes all database character sets and supports the Unicode 5.0 standard. For **WORLD_LEXER** to be effective with documents that use multiple languages, AL32UTF-8 or UTF8 Oracle character set encoding must be specified. This includes supplementary, or "surrogate-pair," characters.

Table D–2 and Table D–3 show the languages supported by **WORLD_LEXER**. This list may change as the Unicode standard changes, and in any case should not be
considered exhaustive. (Languages are grouped by Unicode writing system, not by natural language groupings.)

Table D–2  Languages Supported by the World Lexer (Space-separated)

<table>
<thead>
<tr>
<th>Language Group</th>
<th>Languages Include</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic</td>
<td>Arabic, Farsi, Kurdish, Pashto, Sindhi, Urdu</td>
</tr>
<tr>
<td>Armenian</td>
<td>Armenian</td>
</tr>
<tr>
<td>Bengali</td>
<td>Assamese, Bengali</td>
</tr>
<tr>
<td>Bopomofo</td>
<td>Hakka Chinese, Minnan Chinese</td>
</tr>
<tr>
<td>Cyrillic</td>
<td>Over 50 languages, including Belorussian, Bulgarian, Macedonian, Moldavian, Russian, Serbian, Serbo-Croatian, Ukrainian</td>
</tr>
<tr>
<td>Devenagari</td>
<td>Bhojpuri, Bihari, Hindi, Kashmiri, Marathi, Nepali, Pali, Sanskrit</td>
</tr>
<tr>
<td>Ethiopic</td>
<td>Amharic, Ge’ez, Tigrinya, Tigre</td>
</tr>
<tr>
<td>Georgian</td>
<td>Georgian</td>
</tr>
<tr>
<td>Greek</td>
<td>Greek</td>
</tr>
<tr>
<td>Gujarati</td>
<td>Gujarati, Kacchi</td>
</tr>
<tr>
<td>Gurmukhi</td>
<td>Punjabi</td>
</tr>
<tr>
<td>Hebrew</td>
<td>Hebrew, Ladino, Yiddish</td>
</tr>
<tr>
<td>Kaganga</td>
<td>Redjang</td>
</tr>
<tr>
<td>Kannada</td>
<td>Kanarese, Kannada</td>
</tr>
<tr>
<td>Korean</td>
<td>Korean, Hanja Hangul</td>
</tr>
<tr>
<td>Latin</td>
<td>Afrikaans, Albanian, Basque, Breton, Catalan, Croatian, Czech, Danish, Dutch, English, Esperanto, Estonian, Faeroese, Fijian, Finnish, Flemish, French, Frisian, German, Hawaiian, Hungarian, Icelandic, Indonesian, Irish, Italian, Lappish, Classic Latin, Latvian, Lithuanian, Malay, Maltese, Pinyin Mandarin, Maori, Norwegian, Polish, Portuguese, Provencal, Romanian, Rumanian, Samoan, Scottish Gaelic, Slovak, Slovene, Slovenian, Sorbian, Spanish, Swahili, Swedish, Tagalog, Turkish, Vietnamese, Welsh</td>
</tr>
<tr>
<td>Malayalam</td>
<td>Malayalam</td>
</tr>
<tr>
<td>Mongolian</td>
<td>Mongolian</td>
</tr>
<tr>
<td>Oriya</td>
<td>Oriya</td>
</tr>
<tr>
<td>Sinhalese, Sinhala</td>
<td>Pali, Sinhalese</td>
</tr>
<tr>
<td>Syriac</td>
<td>Aramaic, Syriac</td>
</tr>
<tr>
<td>Tamil</td>
<td>Tamil</td>
</tr>
<tr>
<td>Telugu</td>
<td>Telugu</td>
</tr>
<tr>
<td>Thaana</td>
<td>Dhivehi, Divehi, Maldivian</td>
</tr>
</tbody>
</table>

Table D–3  Languages Supported by the World Lexer (Non-space-separated)

<table>
<thead>
<tr>
<th>Language Group</th>
<th>Languages Include</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>Cantonese, Mandarin, Pinyin phonograms</td>
</tr>
<tr>
<td>Japanese</td>
<td>Japanese (Hiragana, Kanji, Katakana)</td>
</tr>
<tr>
<td>Khmer</td>
<td>Cambodian, Khmer</td>
</tr>
</tbody>
</table>
Querying

Oracle Text supports the use of different query operators. Some operators can be set to behave in accordance with your language. This section summarizes the multilingual query features for these operators.

**ABOUT Operator**

Use the `ABOUT` operator to query on concepts. The system looks up concept information in the theme component of the index.

This feature is supported for English and French with `CONTEXT` indexes only.

**Fuzzy Operator**

This operator enables you to search for words that have similar spelling to specified word. Oracle Text supports `fuzzy` for English, French, German, Italian, Dutch, Spanish, Portuguese, Japanese, Optical Character recognition (OCR), and automatic language detection.

**Stem Operator**

This operator enables you to search for words that have the same root as the specified term. For example, a stem of `$sing` expands into a query on the words `sang`, `sung`, `sing`.

---

**Table D–3 (Cont.) Languages Supported by the World Lexer (Non-space-separated)**

<table>
<thead>
<tr>
<th>Language Group</th>
<th>Languages Include</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buhid</td>
<td>Buhid</td>
</tr>
<tr>
<td>Canadian Syllabics</td>
<td>Blackfoot, Carrier, Cree, Dakhelh, Inuit, Inuktitut, Naskapi,</td>
</tr>
<tr>
<td></td>
<td>Nunavik, Nunavut, Ojibwe, Sayisi, Slavey</td>
</tr>
<tr>
<td>Cherokee</td>
<td>Cherokee</td>
</tr>
<tr>
<td>Cypriot</td>
<td>Cypriot</td>
</tr>
<tr>
<td>Limbu</td>
<td>Limbu</td>
</tr>
<tr>
<td>Ogham</td>
<td>Ogham</td>
</tr>
<tr>
<td>Runic</td>
<td>Runic</td>
</tr>
<tr>
<td>Tai Le (Tai Lu, Lue, Dai Le)</td>
<td>Tai Le</td>
</tr>
<tr>
<td>Ugaritic</td>
<td>Ugaritic</td>
</tr>
<tr>
<td>Yi</td>
<td>Yi</td>
</tr>
<tr>
<td>Yi Jang Hexagram</td>
<td>Yi Jang</td>
</tr>
</tbody>
</table>

**Table D–4 Languages Not Supported by the World Lexer**

<table>
<thead>
<tr>
<th>Language Group</th>
<th>Languages Include</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lao</td>
<td>Lao</td>
</tr>
<tr>
<td>Myanmar</td>
<td>Burmese</td>
</tr>
<tr>
<td>Thai</td>
<td>Thai</td>
</tr>
<tr>
<td>Tibetan</td>
<td>Dzongkha, Tibetan</td>
</tr>
</tbody>
</table>

Table D–4 shows languages not supported by the World Lexer.
The Oracle Text stemmer supports the following languages: English, French, Spanish, Italian, German, Japanese and Dutch.

**Supplied Stop Lists**

A stoplist is a list of words that do not get indexed. These are usually common words in a language such as *this, that,* and *can* in English.

Oracle Text provides a default stoplist for English, Chinese (traditional and simplified), Danish, Dutch, Finnish, French, German, Italian, Portuguese, Spanish, and Swedish. Appendix E, "Oracle Text Supplied Stoplists", lists the stoplists for various languages.

**Knowledge Base**

An Oracle Text knowledge base is a hierarchical tree of concepts used for theme indexing, ABOUT queries, and deriving themes for document services.

Oracle Text supplies knowledge bases in English and French only.

**Knowledge Base Extension**

Extend theme functionality to languages other than English or French by loading your own knowledge base for any single byte white space delimited language, including Spanish.

**Multilingual Features Matrix**

The following table summarizes the multilingual features for the supported languages. Note that the Auto Detect column lists languages that can be automatically detected by AUTO_LEXER.

<table>
<thead>
<tr>
<th>LANGUAGE</th>
<th>AUTO DETECT</th>
<th>ALTERNATE SPELLING</th>
<th>FUZZY MATCHING</th>
<th>LANGUAGE SPECIFIC LEXER</th>
<th>DEFAULT STOP LIST</th>
<th>STEMMING</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENGLISH</td>
<td>Yes</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>GERMAN</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>JAPANESE</td>
<td>Yes</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>FRENCH</td>
<td>Yes</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SPANISH</td>
<td>Yes</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ITALIAN</td>
<td>Yes</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>DUTCH</td>
<td>Yes</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>PORTUGUESE</td>
<td>Yes</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>KOREAN</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>SIMPLIFIED CHINESE</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>TRADITIONAL CHINESE</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>DANISH</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
### Table D–5 (Cont.) Multilingual Features for Supported Languages

<table>
<thead>
<tr>
<th>LANGUAGE</th>
<th>AUTO DETECT</th>
<th>ALTERNATE SPELLING</th>
<th>FUZZY MATCHING</th>
<th>LANGUAGE SPECIFIC LEXER</th>
<th>DEFAULT STOP LIST</th>
<th>STEMMING</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWEDISH</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>FINNISH</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>ARABIC</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>GREEK</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>BOKMAL</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>POLISH</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>RUSSIAN</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>SLOVENIAN</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>THAI</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>CATALAN</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>CROATIAN</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>HEBREW</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>NYNORSK</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>SERBIAN</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>TURKISH</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>CZECH</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>HUNGARIAN</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>PERSIAN</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>SLOVAK</td>
<td>Yes</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Oracle Text Supplied Stoplists

This appendix describes the default stoplists for all the different languages supported and lists the stopwords in each. The following stoplists are described:

- English Default Stoplist
- Chinese Stoplist (Traditional)
- Chinese Stoplist (Simplified)
- Danish (dk) Default Stoplist
- Dutch (nl) Default Stoplist
- Finnish (sf) Default Stoplist
- French (f) Default Stoplist
- German (d) Default Stoplist
- Italian (i) Default Stoplist
- Portuguese (pt) Default Stoplist
- Spanish (e) Default Stoplist
- Swedish (s) Default Stoplist

English Default Stoplist

The following English words are defined as stop words:

<table>
<thead>
<tr>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>did</td>
<td>in</td>
<td>only</td>
<td>then</td>
<td>where</td>
</tr>
<tr>
<td>all</td>
<td>do</td>
<td>into</td>
<td>onto</td>
<td>there</td>
<td>whether</td>
</tr>
<tr>
<td>almost</td>
<td>does</td>
<td>is</td>
<td>or</td>
<td>therefore</td>
<td>which</td>
</tr>
<tr>
<td>also</td>
<td>either</td>
<td>it</td>
<td>our</td>
<td>these</td>
<td>while</td>
</tr>
<tr>
<td>although</td>
<td>for</td>
<td>its</td>
<td>ours</td>
<td>they</td>
<td>who</td>
</tr>
<tr>
<td>an</td>
<td>from</td>
<td>just</td>
<td>s</td>
<td>this</td>
<td>whose</td>
</tr>
<tr>
<td>and</td>
<td>had</td>
<td>ll</td>
<td>shall</td>
<td>those</td>
<td>why</td>
</tr>
<tr>
<td>any</td>
<td>has</td>
<td>me</td>
<td>she</td>
<td>though</td>
<td>will</td>
</tr>
<tr>
<td>are</td>
<td>have</td>
<td>might</td>
<td>should</td>
<td>through</td>
<td>with</td>
</tr>
<tr>
<td>as</td>
<td>having</td>
<td>Mr</td>
<td>since</td>
<td>thus</td>
<td>would</td>
</tr>
</tbody>
</table>
The following traditional Chinese words are defined in the default stoplist for this language.

<table>
<thead>
<tr>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
</tr>
</thead>
<tbody>
<tr>
<td>目前</td>
<td>由於</td>
<td>因此</td>
<td>他們</td>
<td>可能</td>
<td>沒有</td>
</tr>
<tr>
<td>有關</td>
<td>不過</td>
<td>可以</td>
<td>如果</td>
<td>對於</td>
<td>因為</td>
</tr>
<tr>
<td>但是</td>
<td>相當</td>
<td>其中</td>
<td>其他</td>
<td>雖然</td>
<td>我們</td>
</tr>
<tr>
<td>必須</td>
<td>以上</td>
<td>之後</td>
<td>所以</td>
<td>以及</td>
<td>許多</td>
</tr>
<tr>
<td>至於</td>
<td>一般</td>
<td>不是</td>
<td>不能</td>
<td>而且</td>
<td>引起</td>
</tr>
<tr>
<td>除了</td>
<td>不少</td>
<td>最後</td>
<td>就是</td>
<td>分別</td>
<td>加強</td>
</tr>
<tr>
<td>繼續</td>
<td>另外</td>
<td>共同</td>
<td>只有</td>
<td>了解</td>
<td>根據</td>
</tr>
<tr>
<td>過去</td>
<td>所有</td>
<td>不會</td>
<td>以來</td>
<td>任何</td>
<td>一直</td>
</tr>
<tr>
<td>立即</td>
<td>左右</td>
<td>經過</td>
<td>尤其</td>
<td>使得</td>
<td>相關</td>
</tr>
<tr>
<td>進入</td>
<td>並不</td>
<td>據了解</td>
<td>現在</td>
<td>只是</td>
<td>需要</td>
</tr>
<tr>
<td>只要</td>
<td>否則</td>
<td>並未</td>
<td>什麼</td>
<td>如此</td>
<td>不要</td>
</tr>
</tbody>
</table>

Chinese Stoplist (Simplified)

The following simplified Chinese words are defined in the default stoplist for this language.
### Danish (dk) Default Stoplist

The following Danish words are defined in the default stoplist for this language:

<table>
<thead>
<tr>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
</tr>
</thead>
<tbody>
<tr>
<td>af</td>
<td>en</td>
<td>god</td>
<td>hvordan</td>
<td>med</td>
<td>og</td>
<td>udenfor</td>
</tr>
<tr>
<td>aldrig</td>
<td>et</td>
<td>han</td>
<td>I</td>
<td>meget</td>
<td>oppe</td>
<td>under</td>
</tr>
<tr>
<td>alle</td>
<td>endnu</td>
<td>her</td>
<td>De</td>
<td>mellem</td>
<td>på</td>
<td>ved</td>
</tr>
<tr>
<td>altid</td>
<td>få</td>
<td>hos</td>
<td>i</td>
<td>mere</td>
<td>rask</td>
<td>vi</td>
</tr>
<tr>
<td>bagved</td>
<td>lidt</td>
<td>hovfor</td>
<td>imod</td>
<td>mindre</td>
<td>hurtig</td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>fjern</td>
<td>hun</td>
<td>ja</td>
<td>nær</td>
<td>sammen</td>
<td></td>
</tr>
<tr>
<td>der</td>
<td>for</td>
<td>hvad</td>
<td>jeg</td>
<td>hvonår</td>
<td>temmelig</td>
<td></td>
</tr>
<tr>
<td>du</td>
<td>foran</td>
<td>hvem</td>
<td>langsom</td>
<td>nede</td>
<td>nok</td>
<td></td>
</tr>
<tr>
<td>efter</td>
<td>fra</td>
<td>hvor</td>
<td>mange</td>
<td>nej</td>
<td>til</td>
<td></td>
</tr>
<tr>
<td>eller</td>
<td>gennem</td>
<td>hvorhen</td>
<td>måske</td>
<td>nu</td>
<td>uden</td>
<td></td>
</tr>
</tbody>
</table>

### Dutch (nl) Default Stoplist

The following Dutch words are defined in the default stoplist for this language:

<table>
<thead>
<tr>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
</tr>
</thead>
<tbody>
<tr>
<td>aan</td>
<td>betreffende</td>
<td>eer</td>
<td>had</td>
<td>juist</td>
<td>na</td>
<td>overeind</td>
<td>van</td>
<td>weer</td>
<td></td>
</tr>
<tr>
<td>aangaande</td>
<td>bij</td>
<td>eerdat</td>
<td>hadden</td>
<td>jullie</td>
<td>naar</td>
<td>overigens</td>
<td>vandaan</td>
<td>weg</td>
<td></td>
</tr>
<tr>
<td>aangezien</td>
<td>binnen</td>
<td>eerder</td>
<td>hare</td>
<td>kan</td>
<td>nadat</td>
<td>pas</td>
<td>vanuit</td>
<td>wegens</td>
<td></td>
</tr>
<tr>
<td>achter</td>
<td>binnenin</td>
<td>eerlang</td>
<td>heb</td>
<td>klaar</td>
<td>net</td>
<td>precies</td>
<td>vanwege</td>
<td>wel</td>
<td></td>
</tr>
<tr>
<td>achterna</td>
<td>boven</td>
<td>eerst</td>
<td>hebben</td>
<td>kon</td>
<td>niet</td>
<td>reeds</td>
<td>veeleer</td>
<td>weldra</td>
<td></td>
</tr>
<tr>
<td>afgelopen</td>
<td>bovenal</td>
<td>elk</td>
<td>hebt</td>
<td>konden</td>
<td>noch</td>
<td>rond</td>
<td>verder</td>
<td>welk</td>
<td></td>
</tr>
<tr>
<td>al</td>
<td>bovendien</td>
<td>elke</td>
<td>heeft</td>
<td>krachtens</td>
<td>nog</td>
<td>rondom</td>
<td>vervolgens</td>
<td>welke</td>
<td></td>
</tr>
<tr>
<td>aldaar</td>
<td>bovengenoemd</td>
<td>en</td>
<td>hem</td>
<td>kunnen</td>
<td>nogal</td>
<td>sedert</td>
<td>vol</td>
<td>wie</td>
<td></td>
</tr>
<tr>
<td>aldus</td>
<td>bovenstaand</td>
<td>enig</td>
<td>hen</td>
<td>kunt</td>
<td>nu</td>
<td>sinds</td>
<td>volgens</td>
<td>wiens</td>
<td></td>
</tr>
<tr>
<td>alhoewel</td>
<td>bovenvermeld</td>
<td>enigszins</td>
<td>het</td>
<td>later</td>
<td>of</td>
<td>sindsdien</td>
<td>voor</td>
<td>wier</td>
<td></td>
</tr>
<tr>
<td>alias</td>
<td>buiten</td>
<td>enkel</td>
<td>hierbeneden</td>
<td>liever</td>
<td>ofschoon</td>
<td>slechts</td>
<td>vooraf</td>
<td>wij</td>
<td></td>
</tr>
<tr>
<td>alle</td>
<td>daar</td>
<td>er</td>
<td>hierboven</td>
<td>maar</td>
<td>om</td>
<td>sommige</td>
<td>vooral</td>
<td>wijzelf</td>
<td></td>
</tr>
</tbody>
</table>
### Finnish (sf) Default Stoplist

The following Finnish words are defined in the default stoplist for this language:

<table>
<thead>
<tr>
<th>Stopword 1</th>
<th>Stopword 2</th>
<th>Stopword 3</th>
<th>Stopword 4</th>
<th>Stopword 5</th>
<th>Stopword 6</th>
<th>Stopword 7</th>
<th>Stopword 8</th>
<th>Stopword 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>ään</td>
<td>jälkeen</td>
<td>kumpi</td>
<td>nopeasti</td>
<td>suoraan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ah</td>
<td>jo</td>
<td>kumpikaan</td>
<td>nuo</td>
<td>ta</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>joka</td>
<td>kumpikin</td>
<td>nyt</td>
<td>tä</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aina</td>
<td>jokainen</td>
<td>kun</td>
<td>oi</td>
<td>tähdien</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alla</td>
<td>joku</td>
<td>kunhan</td>
<td>olemme</td>
<td>tahi</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alle</td>
<td>jollei</td>
<td>kunnes</td>
<td>olen</td>
<td>tai</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alta</td>
<td>jolleivat</td>
<td>kuten</td>
<td>olet</td>
<td>taikka</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ansiosta</td>
<td>jollemme</td>
<td>kyllä</td>
<td>olette</td>
<td>takana</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>edessä</td>
<td>jollen</td>
<td>kylliksi</td>
<td>oli</td>
<td>takia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>een</td>
<td>jollet</td>
<td>lähellä</td>
<td>olimme</td>
<td>tämä</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ehkä</td>
<td>jollette</td>
<td>läpi</td>
<td>olin</td>
<td>tarpeeksi</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ei</td>
<td>jos</td>
<td>liian</td>
<td>olit</td>
<td>tässä</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eli</td>
<td>joskin</td>
<td>lla</td>
<td>olitte</td>
<td>te</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>elikkä</td>
<td>jotta</td>
<td>llä</td>
<td>olivat</td>
<td>tokko</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ellei</td>
<td>kaikki</td>
<td>lle</td>
<td>ollut</td>
<td>tta</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>elleivät</td>
<td>kanssa</td>
<td>lta</td>
<td>on</td>
<td>ttä</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ellemme</td>
<td>kaukana</td>
<td>ltä</td>
<td>oon</td>
<td>tuo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ellen</td>
<td>ken</td>
<td>luona</td>
<td>ovat</td>
<td>ulkopuolella</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The following French words are defined in the default stoplist for this language:

<table>
<thead>
<tr>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
<th>Stopword</th>
</tr>
</thead>
<tbody>
<tr>
<td>ellet</td>
<td>keneksi</td>
<td>me</td>
<td>paitsi</td>
<td>useammin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ellette</td>
<td>kenelle</td>
<td>mikä</td>
<td>paljon</td>
<td>useimmin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>enemmän</td>
<td>kenkään</td>
<td>mikään</td>
<td>paremmin</td>
<td>usein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eniten</td>
<td>kenties</td>
<td>mikäli</td>
<td>parhaiten</td>
<td>vaan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ennen</td>
<td>keskellä</td>
<td>mikin</td>
<td>pian</td>
<td>vähän</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eräs</td>
<td>kesken</td>
<td>miksi</td>
<td>se</td>
<td>vähemmän</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>että</td>
<td>ketkä</td>
<td>milloin</td>
<td>seen</td>
<td>vähiten</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hän</td>
<td>kohti</td>
<td>milloinkaan</td>
<td>sekä</td>
<td>vaikka</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>harva</td>
<td>koska</td>
<td>minä</td>
<td>sen</td>
<td>vailla</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>he</td>
<td>koskaan</td>
<td>missä</td>
<td>siellä</td>
<td>varten</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hei</td>
<td>ksi</td>
<td>miten</td>
<td>sieltä</td>
<td>vastaan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hitaasti</td>
<td>kuin</td>
<td>molemmat</td>
<td>siin</td>
<td>vielä</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hyi</td>
<td>kuinka</td>
<td>mutta</td>
<td>sillä</td>
<td>voi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hyvin</td>
<td>kuka</td>
<td>na</td>
<td>sinä</td>
<td>ympäri</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ii</td>
<td>kukaan</td>
<td>nää</td>
<td>sinne</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ilman</td>
<td>kukin</td>
<td>nään</td>
<td>ssa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>itse</td>
<td>kumpainen</td>
<td>nämä</td>
<td>ssää</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ja</td>
<td>kumpainenkaa</td>
<td>ne</td>
<td>sta</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>jahka</td>
<td>kumpainenkin</td>
<td>niin</td>
<td>stä</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
German (d) Default Stoplist

The following German words are defined in the default stoplist for this language:

<table>
<thead>
<tr>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
</tr>
</thead>
<tbody>
<tr>
<td>aussitôt</td>
<td>cette</td>
<td>donnée</td>
<td>jadis</td>
<td>mes</td>
<td>où</td>
<td>quiconque</td>
<td>telle</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>autant</td>
<td>ceux</td>
<td>dont</td>
<td>je</td>
<td>mien</td>
<td>par</td>
<td>quoi</td>
<td>telles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>autour</td>
<td>chacun</td>
<td>du</td>
<td>jusqu</td>
<td>mienne</td>
<td>parce</td>
<td>quoique</td>
<td>tes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aux</td>
<td>chacune</td>
<td>duquel</td>
<td>jusque</td>
<td>miennes</td>
<td>parmi</td>
<td>sa</td>
<td>tienne</td>
<td></td>
<td></td>
</tr>
<tr>
<td>auxquelles</td>
<td>chaque</td>
<td>durant</td>
<td>la</td>
<td>miens</td>
<td>plus</td>
<td>sans</td>
<td>tiennes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>auxquels</td>
<td>chez</td>
<td>elle</td>
<td>laquelle</td>
<td>moins</td>
<td>plusieurs</td>
<td>sauf</td>
<td>tiens</td>
<td></td>
<td></td>
</tr>
<tr>
<td>avec</td>
<td>combien</td>
<td>elles</td>
<td>là</td>
<td>moment</td>
<td>pour</td>
<td>se</td>
<td>toi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>à</td>
<td>comme</td>
<td>en</td>
<td>le</td>
<td>mon</td>
<td>pourquoi</td>
<td>selon</td>
<td>ton</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stop word Stop word Stop word Stop word Stop word Stop word Stop word Stop word Stop word Stop word

German (d) Default Stoplist

The following German words are defined in the default stoplist for this language:

<table>
<thead>
<tr>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
<th>German word</th>
</tr>
</thead>
<tbody>
<tr>
<td>ab</td>
<td>dann</td>
<td>des</td>
<td>es</td>
<td>ihnen</td>
<td>keinem</td>
<td>obgleich</td>
<td>sondern</td>
<td>welchem</td>
<td></td>
</tr>
<tr>
<td>aber</td>
<td>daran</td>
<td>desselben</td>
<td>etwa</td>
<td>ihr</td>
<td>keinen</td>
<td>oder</td>
<td>sonst</td>
<td>welchen</td>
<td></td>
</tr>
<tr>
<td>allein</td>
<td>darauf</td>
<td>dessen</td>
<td>etwas</td>
<td>ihre</td>
<td>keiner</td>
<td>ohne</td>
<td>soviel</td>
<td>welcher</td>
<td></td>
</tr>
<tr>
<td>als</td>
<td>daraus</td>
<td>dich</td>
<td>euch</td>
<td>Ihre</td>
<td>keines</td>
<td>paar</td>
<td>soweit</td>
<td>welches</td>
<td></td>
</tr>
<tr>
<td>also</td>
<td>darin</td>
<td>die</td>
<td>euer</td>
<td>ihrem</td>
<td>man</td>
<td>sehr</td>
<td>über</td>
<td>wem</td>
<td></td>
</tr>
<tr>
<td>am</td>
<td>darüber</td>
<td>dies</td>
<td>eure</td>
<td>Ihrem</td>
<td>mehr</td>
<td>sei</td>
<td>um</td>
<td>wen</td>
<td></td>
</tr>
<tr>
<td>an</td>
<td>darum</td>
<td>diese</td>
<td>eurem</td>
<td>ihren</td>
<td>mein</td>
<td>sein</td>
<td>und</td>
<td>wenn</td>
<td></td>
</tr>
<tr>
<td>auch</td>
<td>darunter</td>
<td>dieselbe</td>
<td>euren</td>
<td>Ihnen</td>
<td>meine</td>
<td>seine</td>
<td>uns</td>
<td>wer</td>
<td></td>
</tr>
<tr>
<td>auf</td>
<td>das</td>
<td>dieselben</td>
<td>eurer</td>
<td>Ihrer</td>
<td>meinem</td>
<td>seinem</td>
<td>unser</td>
<td>weshalb</td>
<td></td>
</tr>
<tr>
<td>aus</td>
<td>dasselbe</td>
<td>diesem</td>
<td>eures</td>
<td>ihrer</td>
<td>meinen</td>
<td>seinen</td>
<td>unsre</td>
<td>wessen</td>
<td></td>
</tr>
<tr>
<td>außer</td>
<td>daß</td>
<td>diesen</td>
<td>für</td>
<td>ihres</td>
<td>meiner</td>
<td>seiner</td>
<td>unsrem</td>
<td>wie</td>
<td></td>
</tr>
<tr>
<td>bald</td>
<td>davon</td>
<td>dieser</td>
<td>fürs</td>
<td>ihres</td>
<td>meines</td>
<td>seines</td>
<td>unsren</td>
<td>wir</td>
<td></td>
</tr>
<tr>
<td>bei</td>
<td>davor</td>
<td>dieses</td>
<td>ganz</td>
<td>im</td>
<td>mich</td>
<td>seit</td>
<td>unsrer</td>
<td>wo</td>
<td></td>
</tr>
<tr>
<td>beim</td>
<td>dazu</td>
<td>dir</td>
<td>gar</td>
<td>in</td>
<td>mir</td>
<td>seitdem</td>
<td>unsres</td>
<td>womit</td>
<td></td>
</tr>
<tr>
<td>bin</td>
<td>dazwischen</td>
<td>doch</td>
<td>gegen</td>
<td>ist</td>
<td>mit</td>
<td>selbst</td>
<td>vom</td>
<td>zu</td>
<td></td>
</tr>
<tr>
<td>bis</td>
<td>dein</td>
<td>dorth</td>
<td>genau</td>
<td>ja</td>
<td>nach</td>
<td>sich</td>
<td>von</td>
<td>zum</td>
<td></td>
</tr>
<tr>
<td>bишьchen</td>
<td>deine</td>
<td>du</td>
<td>gewesen</td>
<td>je</td>
<td>nachdem</td>
<td>Sie</td>
<td>vor</td>
<td>zur</td>
<td></td>
</tr>
<tr>
<td>bist</td>
<td>deinem</td>
<td>ebenso</td>
<td>her</td>
<td>jedesmal</td>
<td>nämlich</td>
<td>sie</td>
<td>während</td>
<td>zwar</td>
<td></td>
</tr>
<tr>
<td>da</td>
<td>deinen</td>
<td>ehe</td>
<td>herein</td>
<td>jedoch</td>
<td>neben</td>
<td>sind</td>
<td>war</td>
<td>zwischen</td>
<td></td>
</tr>
<tr>
<td>dabei</td>
<td>deiner</td>
<td>ein</td>
<td>herum</td>
<td>jene</td>
<td>nein</td>
<td>so</td>
<td>wäre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dadurch</td>
<td>deines</td>
<td>eine</td>
<td>hin</td>
<td>jenem</td>
<td>nicht</td>
<td>sogar</td>
<td>wären</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dafür</td>
<td>dem</td>
<td>einem</td>
<td>hinter</td>
<td>jenem</td>
<td>nichts</td>
<td>solch</td>
<td>warum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dagegen</td>
<td>demselben</td>
<td>einen</td>
<td>hintern</td>
<td>jener</td>
<td>noch</td>
<td>solche</td>
<td>was</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dahinter</td>
<td>den</td>
<td>einer</td>
<td>ich</td>
<td>jenes</td>
<td>nun</td>
<td>solchem</td>
<td>wegen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>damit</td>
<td>denn</td>
<td>eines</td>
<td>ihm</td>
<td>kaum</td>
<td>nur</td>
<td>solchen</td>
<td>weil</td>
<td></td>
<td></td>
</tr>
<tr>
<td>danach</td>
<td>der</td>
<td>entlang</td>
<td>ihn</td>
<td>kein</td>
<td>ob</td>
<td>solcher</td>
<td>weit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>daneben</td>
<td>derselben</td>
<td>er</td>
<td>Ihnen</td>
<td>keine</td>
<td>ober</td>
<td>solches</td>
<td>welche</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Italian (i) Default Stoplist

The following Italian words are defined in the default stoplist for this language:

<table>
<thead>
<tr>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>da</td>
<td>durante</td>
<td>lo</td>
<td>o</td>
<td>seppure</td>
<td>un</td>
</tr>
<tr>
<td>affinchè</td>
<td>dachè</td>
<td>e</td>
<td>loro</td>
<td>onde</td>
<td>si</td>
<td>una</td>
</tr>
<tr>
<td>agl’</td>
<td>dagl’</td>
<td>egli</td>
<td>ma</td>
<td>oppure</td>
<td>siccome</td>
<td>uno</td>
</tr>
<tr>
<td>agli</td>
<td>dagli</td>
<td>eppure</td>
<td>mentre</td>
<td>ossia</td>
<td>sopra</td>
<td>voi</td>
</tr>
<tr>
<td>ai</td>
<td>dai</td>
<td>essere</td>
<td>mio</td>
<td>ovvero</td>
<td>sotto</td>
<td>vostro</td>
</tr>
<tr>
<td>al</td>
<td>dal</td>
<td>essi</td>
<td>ne</td>
<td>per</td>
<td>su</td>
<td></td>
</tr>
<tr>
<td>all’</td>
<td>dall’</td>
<td>finché</td>
<td>neanche</td>
<td>perché</td>
<td>subito</td>
<td></td>
</tr>
<tr>
<td>alla</td>
<td>dalla</td>
<td>fino</td>
<td>negl’</td>
<td>percio</td>
<td>sugl’</td>
<td></td>
</tr>
<tr>
<td>alle</td>
<td>dalle</td>
<td>fra</td>
<td>negli</td>
<td>però</td>
<td>sugli</td>
<td></td>
</tr>
<tr>
<td>allo</td>
<td>dallo</td>
<td>giacchè</td>
<td>nei</td>
<td>poichè</td>
<td>suì</td>
<td></td>
</tr>
<tr>
<td>anzichè</td>
<td>degl’</td>
<td>gl’</td>
<td>nel</td>
<td>prima</td>
<td>sul</td>
<td></td>
</tr>
<tr>
<td>avere</td>
<td>degli</td>
<td>gli</td>
<td>nell’</td>
<td>purchè</td>
<td>sull’</td>
<td></td>
</tr>
<tr>
<td>bensì</td>
<td>dei</td>
<td>grazie</td>
<td>nella</td>
<td>quand’anche</td>
<td>sulla</td>
<td></td>
</tr>
<tr>
<td>che</td>
<td>del</td>
<td>l’</td>
<td>nelle</td>
<td>quando</td>
<td>sulle</td>
<td></td>
</tr>
<tr>
<td>chi</td>
<td>dell’</td>
<td>il</td>
<td>nello</td>
<td>quantunque</td>
<td>sullo</td>
<td></td>
</tr>
<tr>
<td>cioè</td>
<td>delle</td>
<td>in</td>
<td>nemmeno</td>
<td>quasi</td>
<td>suo</td>
<td></td>
</tr>
<tr>
<td>come</td>
<td>dello</td>
<td>inoltre</td>
<td>neppure</td>
<td>quindi</td>
<td>talché</td>
<td></td>
</tr>
<tr>
<td>comunque</td>
<td>di</td>
<td>io</td>
<td>noi</td>
<td>se</td>
<td>tu</td>
<td></td>
</tr>
<tr>
<td>con</td>
<td>dopo</td>
<td>l’</td>
<td>nonchè</td>
<td>sebbene</td>
<td>tuo</td>
<td></td>
</tr>
<tr>
<td>contro</td>
<td>dove</td>
<td>la</td>
<td>nondimenno</td>
<td>sennonchè</td>
<td>tuttavia</td>
<td></td>
</tr>
<tr>
<td>cosa</td>
<td>dunque</td>
<td>le</td>
<td>nostro</td>
<td>senza</td>
<td>tutti</td>
<td></td>
</tr>
</tbody>
</table>

Portuguese (pt) Default Stoplist

The following Portuguese words are defined in the default stoplist for this language:

<table>
<thead>
<tr>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>bem</td>
<td>e</td>
<td>longe</td>
<td>para</td>
<td>se</td>
<td>você</td>
</tr>
<tr>
<td>abaixo</td>
<td>com</td>
<td>ela</td>
<td>mais</td>
<td>por</td>
<td>sem</td>
<td>vocês</td>
</tr>
<tr>
<td>adiante</td>
<td>como</td>
<td>elas</td>
<td>menos</td>
<td>porque</td>
<td>sempre</td>
<td></td>
</tr>
<tr>
<td>agora</td>
<td>contra</td>
<td>ele</td>
<td>muito</td>
<td>pouco</td>
<td>sim</td>
<td></td>
</tr>
<tr>
<td>ali</td>
<td>debaixo</td>
<td>eles</td>
<td>não</td>
<td>próximo</td>
<td>sob</td>
<td></td>
</tr>
<tr>
<td>antes</td>
<td>demais</td>
<td>em</td>
<td>ninguem</td>
<td>qual</td>
<td>sobre</td>
<td></td>
</tr>
<tr>
<td>aqui</td>
<td>depois</td>
<td>entre</td>
<td>nós</td>
<td>quando</td>
<td>talvez</td>
<td></td>
</tr>
<tr>
<td>até</td>
<td>depressa</td>
<td>eu</td>
<td>nunca</td>
<td>quanto</td>
<td>todas</td>
<td></td>
</tr>
<tr>
<td>atras</td>
<td>devagar</td>
<td>fora</td>
<td>onde</td>
<td>que</td>
<td>todos</td>
<td></td>
</tr>
<tr>
<td>bastante</td>
<td>direito</td>
<td>junto</td>
<td>ou</td>
<td>quem</td>
<td>vagarosamente</td>
<td></td>
</tr>
</tbody>
</table>

Spanish (e) Default Stoplist

The following Spanish words are defined in the default stoplist for this language:
### Swedish (s) Default Stoplist

The following Swedish words are defined in the default stoplist for this language:

<table>
<thead>
<tr>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
<th>Stop word</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>aquí</td>
<td>cuánto</td>
<td>esta</td>
<td>misma</td>
<td>nosotros</td>
<td>querer</td>
<td>tales</td>
<td>usted</td>
</tr>
<tr>
<td>acá</td>
<td>cada</td>
<td>cuán</td>
<td>estar</td>
<td>mismas</td>
<td>nosotros</td>
<td>qué</td>
<td>tan</td>
<td>ustedes</td>
</tr>
<tr>
<td>ahí</td>
<td>cierta</td>
<td>cuánto</td>
<td>estas</td>
<td>mismo</td>
<td>nuestra</td>
<td>quien</td>
<td>tanta</td>
<td>varias</td>
</tr>
<tr>
<td>ajena</td>
<td>ciertas</td>
<td>cuántos</td>
<td>este</td>
<td>mismos</td>
<td>nuestras</td>
<td>quienes</td>
<td>tantas</td>
<td>varios</td>
</tr>
<tr>
<td>ajenas</td>
<td>cierto</td>
<td>de</td>
<td>estos</td>
<td>mucha</td>
<td>nuestro</td>
<td>quienesera</td>
<td>tanto</td>
<td>vosotras</td>
</tr>
<tr>
<td>ajeno</td>
<td>ciertos</td>
<td>dejar</td>
<td>hacer</td>
<td>muchas</td>
<td>nuestros</td>
<td>quienquiera</td>
<td>tantos</td>
<td>vosotros</td>
</tr>
<tr>
<td>ajenos</td>
<td>como</td>
<td>del</td>
<td>hasta</td>
<td>muchísimas</td>
<td>nunca</td>
<td>quién</td>
<td>te</td>
<td>vuestra</td>
</tr>
<tr>
<td>al</td>
<td>cómo</td>
<td>demasiada</td>
<td>jamás</td>
<td>muchísimas</td>
<td>os</td>
<td>ser</td>
<td>tener</td>
<td>vuestras</td>
</tr>
<tr>
<td>algo</td>
<td>con</td>
<td>demasiado</td>
<td>juntos</td>
<td>muchísimos</td>
<td>otras</td>
<td>siempre</td>
<td>toda</td>
<td>vueistros</td>
</tr>
<tr>
<td>alguna</td>
<td>conmigo</td>
<td>demasiado</td>
<td>la</td>
<td>mucho</td>
<td>otro</td>
<td>sí</td>
<td>todas</td>
<td>y</td>
</tr>
<tr>
<td>algunos</td>
<td>contigo</td>
<td>demasiados</td>
<td>la</td>
<td>muchos</td>
<td>otros</td>
<td>sin</td>
<td>todo</td>
<td>yo</td>
</tr>
<tr>
<td>alguno</td>
<td>cualquier</td>
<td>el</td>
<td>lo</td>
<td>muy</td>
<td>para</td>
<td>Sr</td>
<td>todos</td>
<td></td>
</tr>
<tr>
<td>algún</td>
<td>cualquiera</td>
<td>ella</td>
<td>los</td>
<td>nada</td>
<td>parecer</td>
<td>Sra</td>
<td>tomar</td>
<td></td>
</tr>
<tr>
<td>allá</td>
<td>cualquieras</td>
<td>ellas</td>
<td>mas</td>
<td>ni</td>
<td>poca</td>
<td>Sres</td>
<td>tuya</td>
<td></td>
</tr>
<tr>
<td>allí</td>
<td>cuan</td>
<td>ellos</td>
<td>más</td>
<td>ninguna</td>
<td>pocas</td>
<td>Sta</td>
<td>tuyo</td>
<td></td>
</tr>
<tr>
<td>aquel</td>
<td>cuanta</td>
<td>él</td>
<td>me</td>
<td>ningunas</td>
<td>poco</td>
<td>suya</td>
<td>tú</td>
<td></td>
</tr>
<tr>
<td>aquella</td>
<td>cuantas</td>
<td>esa</td>
<td>menos</td>
<td>ninguno</td>
<td>pocos</td>
<td>suyas</td>
<td>un</td>
<td></td>
</tr>
<tr>
<td>aquellas</td>
<td>cuánta</td>
<td>esas</td>
<td>más</td>
<td>ningunos</td>
<td>por</td>
<td>suyo</td>
<td>una</td>
<td></td>
</tr>
<tr>
<td>aquello</td>
<td>cuántas</td>
<td>ese</td>
<td>mientras</td>
<td>no</td>
<td>porque</td>
<td>suyos</td>
<td>unas</td>
<td></td>
</tr>
<tr>
<td>aquellos</td>
<td>cuanto</td>
<td>esos</td>
<td>mio</td>
<td>nos</td>
<td>que</td>
<td>tal</td>
<td>unos</td>
<td></td>
</tr>
<tr>
<td>Stopword</td>
<td>Stopword</td>
<td>Stopword</td>
<td>Stopword</td>
<td>Stopword</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>åter</td>
<td>fall</td>
<td>inte</td>
<td>ni</td>
<td>upp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>att</td>
<td>färre</td>
<td>ja</td>
<td>nu</td>
<td>uppe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>av</td>
<td>fastän</td>
<td>jag</td>
<td>och</td>
<td>ut</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avse</td>
<td>flest</td>
<td>kan</td>
<td>också</td>
<td>utan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avsedd</td>
<td>flesta</td>
<td>kort</td>
<td>om</td>
<td>ute</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avsedda</td>
<td>för</td>
<td>korta</td>
<td>oss</td>
<td>utom</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avser</td>
<td>först</td>
<td>kunde</td>
<td>över</td>
<td>vad</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avses</td>
<td>första</td>
<td>kunna</td>
<td>överst</td>
<td>väl</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bakom</td>
<td>förste</td>
<td>lång</td>
<td>översta</td>
<td>var</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bara</td>
<td>fort</td>
<td>långa</td>
<td>övre</td>
<td>vara</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bäst</td>
<td>framför</td>
<td>långsam</td>
<td>på</td>
<td>varför</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bättre</td>
<td>från</td>
<td>långsamma</td>
<td>så</td>
<td>vart</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bra</td>
<td>genom</td>
<td>långsamt</td>
<td>sådan</td>
<td>vem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bredvid</td>
<td>god</td>
<td>långt</td>
<td>sådana</td>
<td>vems</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>då</td>
<td>goda</td>
<td>lite</td>
<td>sådant</td>
<td>vet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dålig</td>
<td>gott</td>
<td>liten</td>
<td>säga</td>
<td>veta</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>där</td>
<td>ha</td>
<td>litet</td>
<td>säger</td>
<td>vi</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>därför</td>
<td>hade</td>
<td>man</td>
<td>sägs</td>
<td>vid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>haft</td>
<td>med</td>
<td>sämre</td>
<td>vilken</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dem</td>
<td>han</td>
<td>medan</td>
<td>sämst</td>
<td>vill</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>den</td>
<td>hans</td>
<td>mellan</td>
<td>sân</td>
<td>ville</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>denna</td>
<td>här</td>
<td>men</td>
<td>sånt</td>
<td>visste</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>deras</td>
<td>hellre</td>
<td>mer</td>
<td>såsom</td>
<td>vore</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dess</td>
<td>henne</td>
<td>mera</td>
<td>sin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dessa</td>
<td>hennes</td>
<td>mest</td>
<td>sist</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>det</td>
<td>heta</td>
<td>mesta</td>
<td>sista</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>detta</td>
<td>heter</td>
<td>mindre</td>
<td>ska</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Oracle Text Scoring Algorithm

This appendix describes how Oracle Text calculates scoring for word queries, which is different from the way it calculates scores for ABOUT queries in English. Scoring is obtained using the SCORE operator.

This appendix contains these topics:

- Scoring Algorithm for Word Queries
- Word Scoring Example
- DML and Scoring Algorithm

See Also: "DEFINESCORE" on page 3-13 and "DEFINEMERGE" on page 3-12 for information about user-defined scoring

Scoring Algorithm for Word Queries

To calculate a relevance score for a returned document in a word query, Oracle Text uses an inverse frequency algorithm based on Salton's formula.

Inverse frequency scoring assumes that frequently occurring terms in a document set are noise terms, and so these terms are scored lower. For a document to score high, the query term must occur frequently in the document but infrequently in the document set as a whole.

The following table illustrates Oracle Text's inverse frequency scoring. The first column shows the number of documents in the document set, and the second column shows the number of terms in the document necessary to score 100.

This table assumes that only one document in the set contains the query term.

<table>
<thead>
<tr>
<th>Number of Documents in Document Set</th>
<th>Occurrences of Term in Document Needed to Score 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>34</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>50</td>
<td>13</td>
</tr>
<tr>
<td>100</td>
<td>12</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
</tr>
<tr>
<td>1,000</td>
<td>9</td>
</tr>
<tr>
<td>10,000</td>
<td>7</td>
</tr>
</tbody>
</table>
### Scoring Algorithm for Word Queries

<table>
<thead>
<tr>
<th>Number of Documents in Document Set</th>
<th>Occurrences of Term in Document Needed to Score 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>100,000</td>
<td>6</td>
</tr>
<tr>
<td>1,000,000</td>
<td>5</td>
</tr>
</tbody>
</table>

Note that the score varies, depending on the set size. For example, if only one document in the set contains the query term, and there are five documents in the set, then the term must occur 20 times in the document to score 100. If 1,000,000 documents are in the set, then the term can occur only 5 times in the document to score 100.

### Word Scoring Example

You have 5000 documents dealing with chemistry in which the term chemical occurs at least once in every document. The term chemical thus occurs frequently in the document set.

You have a document that contains 5 occurrences of chemical and 5 occurrences of the term hydrogen. No other document contains the term hydrogen. The term hydrogen thus occurs infrequently in the document set.

Because chemical occurs so frequently in the document set, its score for the document is lower with respect to hydrogen, which is infrequent is the document set as a whole. The score for hydrogen is therefore higher than that of chemical. This is so even though both terms occur 5 times in the document.

### Note

Even if the relatively infrequent term hydrogen occurred 4 times in the document, and chemical occurred 5 times in the document, the score for hydrogen might still be higher, because chemical occurs so frequently in the document set (at least 5000 times).

Inverse frequency scoring also means that adding documents that contain hydrogen lowers the score for that term in the document, and adding more documents that do not contain hydrogen raises the score.

### DML and Scoring Algorithm

Because the scoring algorithm is based on the number of documents in the document set, inserting, updating or deleting documents in the document set is likely to change the score for any given term before and after DML.

If DML is heavy, you must optimize the index. Perfect relevance ranking is obtained by running a query right after optimizing the index.

If DML is light, Oracle Database still gives fairly accurate relevance ranking.

In either case, you must synchronize the index with CTX_DDL.SYNC_INDEX.

**See Also:**  "SYNC_INDEX" on page 7-75
This appendix lists all of the views provided by Oracle Text. The system provides the following views:

- CTX_CLASSES
- CTX_EXTRACT_POLICIES
- CTX_EXTRACT_POLICY_VALUES
- CTX_FILTER_BY_COLUMNS
- CTX_INDEXES
- CTX_INDEX_ERRORS
- CTX_INDEX_OBJECTS
- CTX_INDEX_PARTITIONS
- CTX_INDEX_SETS
- CTX_INDEX_SET_INDEXES
- CTX_INDEX_SUB_LEXERS
- CTX_INDEX_SUB_LEXER_VALUES
- CTX_INDEX_VALUES
- CTX_OBJECTS
- CTX_OBJECT_ATTRIBUTES
- CTX_OBJECT_ATTRIBUTE_LOV
- CTX_ORDER_BY_COLUMNS
- CTX_PARAMETERS
- CTX_PENDING
- CTX_PREFERENCES
- CTX_PREFERENCE_VALUES
- CTX_SECTIONS
- CTX_SECTION_GROUPS
- CTX_SQES
- CTX_STOPLISTS
- CTX_STOPWORDS
This view displays all the preference categories registered in the Text data dictionary. It can be queried by any user.
**CTX_EXTRACT_POLICIES**

This view displays all entity extraction policies. Only **CTXSYS** can query this view.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPL_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Name of the entity extraction policy</td>
</tr>
<tr>
<td>EPL_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Owner of the entity extraction policy</td>
</tr>
</tbody>
</table>

**CTX_EXTRACT_POLICY_VALUES**

This view displays the values for all entity extraction policies. Only **CTXSYS** can query this view.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPV_POLICY_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Entity extraction policy name</td>
</tr>
<tr>
<td>EPV_POLICY_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Entity extraction policy owner</td>
</tr>
<tr>
<td>EPV_CLASS</td>
<td>VARCHAR2 (30)</td>
<td>Object class</td>
</tr>
<tr>
<td>EPV_OBJECT</td>
<td>VARCHAR2 (30)</td>
<td>Object name</td>
</tr>
<tr>
<td>EPV_ATTRIBUTE</td>
<td>VARCHAR2 (30)</td>
<td>Object attribute name</td>
</tr>
<tr>
<td>EPV_VALUE</td>
<td>VARCHAR2 (500)</td>
<td>Object attribute value</td>
</tr>
</tbody>
</table>

**CTX_FILTER_BY_COLUMNS**

This view displays all **FILTER BY** columns registered in the Text data dictionary. It can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBC_INDEX_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Index owner name</td>
</tr>
<tr>
<td>FBC_INDEX_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Index name</td>
</tr>
<tr>
<td>FBC_TABLE_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Table owner name</td>
</tr>
<tr>
<td>FBC_TABLE_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Table name</td>
</tr>
<tr>
<td>FBC_COLUMN_NAME</td>
<td>VARCHAR2 (256)</td>
<td>Column name</td>
</tr>
<tr>
<td>FBC_COLUMN_TYPE</td>
<td>VARCHAR2 (30)</td>
<td>Column type</td>
</tr>
<tr>
<td>FBC_SECTION_TYPE</td>
<td>VARCHAR2 (30)</td>
<td>Section type</td>
</tr>
<tr>
<td>FBC_SECTION_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Section name</td>
</tr>
<tr>
<td>FBC_SECTION_ID</td>
<td>NUMBER</td>
<td>Section ID</td>
</tr>
</tbody>
</table>
**CTX_INDEXES**

This view displays all indexes that are registered in the Text data dictionary for the current user. It can be queried by CTXSYS.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDX_CHARSET_COLUMN</td>
<td>VARCHAR2(256)</td>
<td>Name of the charset column in base table.</td>
</tr>
<tr>
<td>IDX_DOCID_COUNT</td>
<td>NUMBER</td>
<td>Number of documents indexed.</td>
</tr>
<tr>
<td>IDX_FORMAT_COLUMNS</td>
<td>VARCHAR2(256)</td>
<td>Name of the format column in base table.</td>
</tr>
<tr>
<td>IDX_ID</td>
<td>NUMBER</td>
<td>Internal index ID.</td>
</tr>
<tr>
<td>IDX_KEY_NAME</td>
<td>VARCHAR2(256)</td>
<td>Primary key column(s).</td>
</tr>
<tr>
<td>IDX_LANGUAGE_COLUMN</td>
<td>VARCHAR2(256)</td>
<td>Name of the language column in base table.</td>
</tr>
<tr>
<td>IDX_NAME</td>
<td>VARCHAR2(30)</td>
<td>Name of index.</td>
</tr>
<tr>
<td>IDX_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Owner of index.</td>
</tr>
<tr>
<td>IDX_STATUS</td>
<td>VARCHAR2(12)</td>
<td>Status.</td>
</tr>
<tr>
<td>IDX_SYNC_JOBNAME</td>
<td>VARCHAR2(50)</td>
<td>Scheduler job name for automatic sync. Only meaningful for AUTOMATIC sync and always null for other types of sync.</td>
</tr>
<tr>
<td>IDX_SYNC_MEMORY</td>
<td>VARCHAR2(100)</td>
<td>Sync memory size. Only meaningful for ON COMMIT and AUTOMATIC types of sync. For MANUAL sync, this is always null.</td>
</tr>
<tr>
<td>IDX_SYNC_PARA_DEGREE</td>
<td>NUMBER</td>
<td>Degree of parallelism for sync. Only meaningful for the AUTOMATIC type of sync; always null for MANUAL and ON COMMIT syncs.</td>
</tr>
<tr>
<td>IDX_SYNC_TYPE</td>
<td>VARCHAR2(20)</td>
<td>Type of synching: MANUAL, AUTOMATIC, or ON COMMIT.</td>
</tr>
<tr>
<td>IDX_TABLE</td>
<td>VARCHAR2(30)</td>
<td>Table name.</td>
</tr>
<tr>
<td>IDX_TABLE_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Owner of table.</td>
</tr>
<tr>
<td>IDX_TEXT_NAME</td>
<td>VARCHAR2(30)</td>
<td>Text column name.</td>
</tr>
<tr>
<td>IDX_TYPE</td>
<td>VARCHAR2(7)</td>
<td>Type of index: CONTEXT, CTXCAT, or CTXRULE.</td>
</tr>
</tbody>
</table>

**CTX_INDEX_ERRORS**

This view displays the DML errors and is queryable by CTXSYS.
**CTX_INDEX_SETS**

This view displays all index set names. It can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERR_INDEX_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Index owner.</td>
</tr>
<tr>
<td>ERR_INDEX_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Name of index.</td>
</tr>
<tr>
<td>ERR_TIMESTAMP</td>
<td>DATE</td>
<td>Time of error.</td>
</tr>
<tr>
<td>ERR_TEXTKEY</td>
<td>VARCHAR2 (18)</td>
<td>ROWID of errored document or name of errored operation (for example, ALTER INDEX)</td>
</tr>
<tr>
<td>ERR_TEXT</td>
<td>VARCHAR2 (4000)</td>
<td>Error text.</td>
</tr>
</tbody>
</table>

**CTX_INDEX_OBJECTS**

This view displays the objects that are used for each class in the index. It can be queried by CTXSYS.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXO_INDEX_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Index owner.</td>
</tr>
<tr>
<td>IXO_INDEX_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Index name.</td>
</tr>
<tr>
<td>IXO_CLASS</td>
<td>VARCHAR2 (30)</td>
<td>Class name.</td>
</tr>
<tr>
<td>IXO_OBJECT</td>
<td>VARCHAR2 (30)</td>
<td>Object name.</td>
</tr>
</tbody>
</table>

**CTX_INDEX_PARTITIONS**

This view displays all index partitions. It can be queried by CTXSYS.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXP_ID</td>
<td>NUMBER (38)</td>
<td>Index partition ID.</td>
</tr>
<tr>
<td>IXP_INDEX_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Index owner.</td>
</tr>
<tr>
<td>IXP_INDEX_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Index name.</td>
</tr>
<tr>
<td>IXP_INDEX_PARTITION_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Index partition name.</td>
</tr>
<tr>
<td>IXP_SYNC_TYPE</td>
<td>VARCHAR2 (20)</td>
<td>Type of synching: MANUAL, AUTOMATIC, or ON COMMIT.</td>
</tr>
<tr>
<td>IXP_TABLE_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Table owner.</td>
</tr>
<tr>
<td>IXP_TABLE_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Table name.</td>
</tr>
<tr>
<td>IXP_TABLE_PARTITION_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Table partition name.</td>
</tr>
<tr>
<td>IXP_DOCID_COUNT</td>
<td>NUMBER (38)</td>
<td>Number of documents associated with the partition.</td>
</tr>
<tr>
<td>IXP_STATUS</td>
<td>VARCHAR2 (12)</td>
<td>Partition status.</td>
</tr>
</tbody>
</table>
**CTX_INDEX_SET_INDEXES**

This view displays all the sub-indexes in an index set. It can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXS_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Index set owner.</td>
</tr>
<tr>
<td>IXS_NAME</td>
<td>VARCHAR2(30)</td>
<td>Index set name.</td>
</tr>
</tbody>
</table>

**CTX_INDEX_SUB_LEXERS**

This view shows the sub-lexers for each language for each index. It can be queried by CTXSYS.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISL_INDEX_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Index owner.</td>
</tr>
<tr>
<td>ISL_INDEX_NAME</td>
<td>VARCHAR2(30)</td>
<td>Index name.</td>
</tr>
<tr>
<td>ISL_LANGUAGE</td>
<td>VARCHAR2(30)</td>
<td>Language of sub-lexer</td>
</tr>
<tr>
<td>ISL_ALT_VALUE</td>
<td>VARCHAR2(30)</td>
<td>Alternate value of language</td>
</tr>
<tr>
<td>ISL_OBJECT</td>
<td>VARCHAR2(30)</td>
<td>Name of lexer object used for this language.</td>
</tr>
</tbody>
</table>

**CTX_INDEX_SUB_LEXER_VALUES**

Shows the sub-lexer attributes and their values. Accessible by CTXSYS.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISV_INDEX_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Index owner.</td>
</tr>
<tr>
<td>ISV_INDEX_NAME</td>
<td>VARCHAR2(30)</td>
<td>Index name.</td>
</tr>
<tr>
<td>ISV_LANGUAGE</td>
<td>VARCHAR2(30)</td>
<td>Language of sub-lexer</td>
</tr>
<tr>
<td>ISV_OBJECT</td>
<td>VARCHAR2(30)</td>
<td>Name of lexer object used for this language.</td>
</tr>
<tr>
<td>ISV_ATTRIBUTE</td>
<td>VARCHAR2(30)</td>
<td>Name of sub-lexer attribute</td>
</tr>
<tr>
<td>ISV_VALUE</td>
<td>VARCHAR2(500)</td>
<td>Value of attribute of sub-lexer</td>
</tr>
</tbody>
</table>

**CTX_INDEX_VALUES**

This view displays attribute values for each object used in indexes. This view is queryable by CTXSYS.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISV_INDEX_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Index owner.</td>
</tr>
<tr>
<td>ISV_INDEX_NAME</td>
<td>VARCHAR2(30)</td>
<td>Index name.</td>
</tr>
<tr>
<td>ISV_LANGUAGE</td>
<td>VARCHAR2(30)</td>
<td>Language of sub-lexer</td>
</tr>
<tr>
<td>ISV_OBJECT</td>
<td>VARCHAR2(30)</td>
<td>Name of lexer object used for this language.</td>
</tr>
<tr>
<td>ISV_ATTRIBUTE</td>
<td>VARCHAR2(30)</td>
<td>Name of sub-lexer attribute</td>
</tr>
<tr>
<td>ISV_VALUE</td>
<td>VARCHAR2(500)</td>
<td>Value of attribute of sub-lexer</td>
</tr>
</tbody>
</table>
**CTX_OBJECTS**

This view displays all of the Text objects registered in the Text data dictionary. This view can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXV_INDEX_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Index owner.</td>
</tr>
<tr>
<td>IXV_INDEX_NAME</td>
<td>VARCHAR2(30)</td>
<td>Index name.</td>
</tr>
<tr>
<td>IXV_CLASS</td>
<td>VARCHAR2(30)</td>
<td>Class name.</td>
</tr>
<tr>
<td>IXV_OBJECT</td>
<td>VARCHAR2(30)</td>
<td>Object name.</td>
</tr>
<tr>
<td>IXV_ATTRIBUTE</td>
<td>VARCHAR2(30)</td>
<td>Attribute name.</td>
</tr>
<tr>
<td>IXV_VALUE</td>
<td>VARCHAR2(500)</td>
<td>Attribute value.</td>
</tr>
</tbody>
</table>

**CTX_OBJECT_ATTRIBUTES**

This view displays the attributes that can be assigned to preferences of each object. It can be queried by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBJ_CLASS</td>
<td>VARCHAR2(30)</td>
<td>Object class (Datastore, Filter, Lexer, and so on)</td>
</tr>
<tr>
<td>OBJ_NAME</td>
<td>VARCHAR2(30)</td>
<td>Object name</td>
</tr>
<tr>
<td>OBJ_DESCRIPTION</td>
<td>VARCHAR2(80)</td>
<td>Object description</td>
</tr>
</tbody>
</table>

**CTX_OBJECT_ATTRIBUTE_LOV**

This view displays the allowed values for certain object attributes provided by Oracle Text. It can be queried by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXV_INDEX_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Index owner.</td>
</tr>
<tr>
<td>IXV_INDEX_NAME</td>
<td>VARCHAR2(30)</td>
<td>Index name.</td>
</tr>
<tr>
<td>IXV_CLASS</td>
<td>VARCHAR2(30)</td>
<td>Class name.</td>
</tr>
<tr>
<td>IXV_OBJECT</td>
<td>VARCHAR2(30)</td>
<td>Object name.</td>
</tr>
<tr>
<td>IXV_ATTRIBUTE</td>
<td>VARCHAR2(30)</td>
<td>Attribute name.</td>
</tr>
<tr>
<td>IXV_VALUE</td>
<td>VARCHAR2(500)</td>
<td>Attribute value.</td>
</tr>
</tbody>
</table>

**Column Name**

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXV_INDEX_OWNER</td>
<td>VARCHAR2(30)</td>
</tr>
<tr>
<td>IXV_INDEX_NAME</td>
<td>VARCHAR2(30)</td>
</tr>
<tr>
<td>IXV_CLASS</td>
<td>VARCHAR2(30)</td>
</tr>
<tr>
<td>IXV_OBJECT</td>
<td>VARCHAR2(30)</td>
</tr>
<tr>
<td>IXV_ATTRIBUTE</td>
<td>VARCHAR2(30)</td>
</tr>
<tr>
<td>IXV_VALUE</td>
<td>VARCHAR2(500)</td>
</tr>
</tbody>
</table>

**CTX_OBJECT_ATTRIBUTE_LOV**

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXV_INDEX_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Index owner.</td>
</tr>
<tr>
<td>IXV_INDEX_NAME</td>
<td>VARCHAR2(30)</td>
<td>Index name.</td>
</tr>
<tr>
<td>IXV_CLASS</td>
<td>VARCHAR2(30)</td>
<td>Class name.</td>
</tr>
<tr>
<td>IXV_OBJECT</td>
<td>VARCHAR2(30)</td>
<td>Object name.</td>
</tr>
<tr>
<td>IXV_ATTRIBUTE</td>
<td>VARCHAR2(30)</td>
<td>Attribute name.</td>
</tr>
<tr>
<td>IXV_VALUE</td>
<td>VARCHAR2(500)</td>
<td>Attribute value.</td>
</tr>
</tbody>
</table>
This view displays the ORDER BY columns registered in the Text data dictionary. It can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAL_CLASS</td>
<td>NUMBER(38)</td>
<td>Class of object.</td>
</tr>
<tr>
<td>OAL_OBJECT</td>
<td>VARCHAR2(30)</td>
<td>Object name.</td>
</tr>
<tr>
<td>OAL_ATTRIBUTE</td>
<td>VARCHAR2(32)</td>
<td>Attribute name.</td>
</tr>
<tr>
<td>OAL_LABEL</td>
<td>VARCHAR2(30)</td>
<td>Attribute value label.</td>
</tr>
<tr>
<td>OAL_VALUE</td>
<td>VARCHAR2(64)</td>
<td>Attribute value.</td>
</tr>
<tr>
<td>OAL_DESCRIPTION</td>
<td>VARCHAR2(80)</td>
<td>Attribute value description.</td>
</tr>
</tbody>
</table>

This view displays all system-defined parameters as defined by CTXSYS. It can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBC_INDEX_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Index owner</td>
</tr>
<tr>
<td>OBC_INDEX_NAME</td>
<td>VARCHAR2(30)</td>
<td>Index name</td>
</tr>
<tr>
<td>OBC_TABLE_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Table owner</td>
</tr>
<tr>
<td>OBC_TABLE_NAME</td>
<td>VARCHAR2(30)</td>
<td>Table name</td>
</tr>
<tr>
<td>OBC_COLUMN_NAME</td>
<td>VARCHAR2(236)</td>
<td>Column name</td>
</tr>
<tr>
<td>OBC_COLUMN_POSITION</td>
<td>VARCHAR2(30)</td>
<td>Column position</td>
</tr>
<tr>
<td>OBC_COLUMN_TYPE</td>
<td>VARCHAR2(30)</td>
<td>Column type</td>
</tr>
<tr>
<td>OBC_SECTION_NAME</td>
<td>VARCHAR2(30)</td>
<td>Section name</td>
</tr>
<tr>
<td>OBC_SECTION_TYPE</td>
<td>VARCHAR2(30)</td>
<td>Section type</td>
</tr>
<tr>
<td>OBC_SECTION_ID</td>
<td>NUMBER</td>
<td>Section ID</td>
</tr>
<tr>
<td>OBC_SORT_ORDER</td>
<td>VARCHAR2(8)</td>
<td>Sort order</td>
</tr>
</tbody>
</table>
This view displays a row for each of the user's entries in the DML Queue. It can be queried by CTXSYS.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAR_NAME</td>
<td>VARCHAR2(30)</td>
<td>Parameter name: max_index_memory, ctx_doc_key_type, default_index_memory</td>
</tr>
<tr>
<td></td>
<td></td>
<td>default_datastore, default_filter_binary, default_filter_text, default_filter_file, default_section_html, default_section_xml, default_section_text, default Lexer, default_stoplist, default_storage, default_wordlist, default_ctxcat Lexer, default_ctxcat_index_set, default_ctxcat_stoplist, default_ctxcat_storage, default_ctxcat_wordlist, default_ctxrule Lexer, default_ctxrule_stoplist, default_ctxrule_storage, default_ctxrule_wordlist, log_directory, file_access_role.</td>
</tr>
<tr>
<td>PAR_VALUE</td>
<td>VARCHAR2(500)</td>
<td>Parameter value. For max_index_memory and default_index_memory, PAR_VALUE stores a string consisting of the memory amount. For the other parameter names, PAR_VALUE stores the names of the preferences used as defaults for index creation.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PND_INDEX_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Index owner.</td>
</tr>
<tr>
<td>PND_INDEX_NAME</td>
<td>VARCHAR2(30)</td>
<td>Name of index.</td>
</tr>
</tbody>
</table>
**CTX_PREFERENCES**

This view displays preferences created by Oracle Text users, as well as all the system-defined preferences included with Oracle Text. The view contains one row for each preference. It can be queried by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PND_PARTITION_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Name of partition for local partition indexes. NULL for normal indexes.</td>
</tr>
<tr>
<td>PND_ROWID</td>
<td>ROWID</td>
<td>ROWID to be indexed</td>
</tr>
<tr>
<td>PND_TIMESTAMP</td>
<td>DATE</td>
<td>Time of modification</td>
</tr>
</tbody>
</table>

**CTX_PREFERENCE_VALUES**

This view displays the values assigned to all the preferences in the Text data dictionary. The view contains one row for each value. It can be queried by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRV_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Username of preference owner.</td>
</tr>
<tr>
<td>PRV_PREFERENCE</td>
<td>VARCHAR2 (30)</td>
<td>Preference name.</td>
</tr>
<tr>
<td>PRV_ATTRIBUTE</td>
<td>VARCHAR2 (64)</td>
<td>Attribute name</td>
</tr>
<tr>
<td>PRV_VALUE</td>
<td>VARCHAR2 (500)</td>
<td>Attribute value</td>
</tr>
</tbody>
</table>

**CTX_SECTIONS**

This view displays information about all the sections that have been created in the Text data dictionary. It can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEC_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Owner of the section group.</td>
</tr>
<tr>
<td>SEC_SECTION_GROUP</td>
<td>VARCHAR2 (30)</td>
<td>Name of the section group.</td>
</tr>
<tr>
<td>SEC_TYPE</td>
<td>VARCHAR2 (30)</td>
<td>Type of section, either ZONE, FIELD, SPECIAL, ATTR, STOP.</td>
</tr>
<tr>
<td>SEC_ID</td>
<td>NUMBER</td>
<td>Section ID.</td>
</tr>
<tr>
<td>SEC_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Name of section.</td>
</tr>
<tr>
<td>SEC_TAG</td>
<td>VARCHAR2 (64)</td>
<td>Section tag</td>
</tr>
<tr>
<td>SEC_VISIBLE</td>
<td>VARCHAR2 (1)</td>
<td>Y or N visible indicator for field sections only.</td>
</tr>
</tbody>
</table>
This view displays information about all the section groups that have been created in the Text data dictionary. It can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGP_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Owner of section group.</td>
</tr>
<tr>
<td>SGP_NAME</td>
<td>VARCHAR2(30)</td>
<td>Name of section group.</td>
</tr>
<tr>
<td>SGP_TYPE</td>
<td>VARCHAR2(30)</td>
<td>Type of section group</td>
</tr>
</tbody>
</table>

This view displays the definitions for all SQEs that have been created by users. It can be queried by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQE_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Owner of SQE.</td>
</tr>
<tr>
<td>SQE_NAME</td>
<td>VARCHAR2(30)</td>
<td>Name of SQE.</td>
</tr>
<tr>
<td>SQE_QUERY</td>
<td>CLOB</td>
<td>Query Text.</td>
</tr>
</tbody>
</table>

This view displays stoplists. Queryable by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPL_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Owner of stoplist.</td>
</tr>
<tr>
<td>SPL_NAME</td>
<td>VARCHAR2(30)</td>
<td>Name of stoplist.</td>
</tr>
<tr>
<td>SPL_COUNT</td>
<td>NUMBER</td>
<td>Number of stopwords</td>
</tr>
<tr>
<td>SPL_TYPE</td>
<td>VARCHAR2(30)</td>
<td>Type of stoplist, MULTI or BASIC.</td>
</tr>
</tbody>
</table>

This view displays the stopwords in each stoplist. Queryable by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPW_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Stoplist owner.</td>
</tr>
<tr>
<td>SPW_STOPLIST</td>
<td>VARCHAR2(30)</td>
<td>Stoplist name.</td>
</tr>
<tr>
<td>SPW_TYPE</td>
<td>VARCHAR2(10)</td>
<td>Stop type, either STOP_WORD, STOP_CLASS, STOP_THEME.</td>
</tr>
<tr>
<td>SPW_WORD</td>
<td>VARCHAR2(80)</td>
<td>Stopword.</td>
</tr>
<tr>
<td>SPW_LANGUAGE</td>
<td>VARCHAR2(30)</td>
<td>Stopword language.</td>
</tr>
</tbody>
</table>
CTX_SUB_LEXERS

This view contains information on multi-lexers and the sub-lexer preferences they contain. It can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLX_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Owner of the multi-lexer preference.</td>
</tr>
<tr>
<td>SLX_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Name of the multi-lexer preference.</td>
</tr>
<tr>
<td>SLX_LANGUAGE</td>
<td>VARCHAR2 (30)</td>
<td>Language of the referenced lexer (full name, not abbreviation).</td>
</tr>
<tr>
<td>SLX_ALT_VALUE</td>
<td>VARCHAR2 (30)</td>
<td>An alternate value for the language.</td>
</tr>
<tr>
<td>SLX_SUB_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Owner of the sub-lexer.</td>
</tr>
<tr>
<td>SLX_SUB_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Name of the sub-lexer.</td>
</tr>
</tbody>
</table>

CTX_ThES_AURI

This view displays information about all the thesauri that have been created in the Text data dictionary. It can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>THS_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Thesaurus owner.</td>
</tr>
<tr>
<td>THS_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Thesaurus name.</td>
</tr>
</tbody>
</table>

CTX_THES_PHRASES

This view displays phrase information for all thesauri in the Text data dictionary. It can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>THP_THESAURUS</td>
<td>VARCHAR2 (30)</td>
<td>Thesaurus name.</td>
</tr>
<tr>
<td>THP_PHRASE</td>
<td>VARCHAR2 (256)</td>
<td>Thesaurus phrase.</td>
</tr>
<tr>
<td>THP_QUALIFIER</td>
<td>VARCHAR2 (256)</td>
<td>Thesaurus qualifier.</td>
</tr>
</tbody>
</table>

CTX_TRACE_VALUES

This view contains one row for each active trace, and shows the current value of each trace.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRC_ID</td>
<td>BINARY_INTEGER</td>
<td>Trace ID.</td>
</tr>
<tr>
<td>TRC_VALUE</td>
<td>NUMBER</td>
<td>Current trace value.</td>
</tr>
</tbody>
</table>

CTX_USER_EXTRACT_POLICIES

This view displays all of the entity extraction policies owned by the current user. All users can query this view.
This view displays all of the values for the entity extraction policies owned by the current user. All users can query this view.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPL_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Entity extraction policy name</td>
</tr>
<tr>
<td>EPV_POLICY_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Entity extraction policy name</td>
</tr>
<tr>
<td>EPV_CLASS</td>
<td>VARCHAR2 (30)</td>
<td>Object class</td>
</tr>
<tr>
<td>EPV_OBJECT</td>
<td>VARCHAR2 (30)</td>
<td>Object name</td>
</tr>
<tr>
<td>EPV_ATTRIBUTE</td>
<td>VARCHAR2 (30)</td>
<td>Object attribute name</td>
</tr>
<tr>
<td>EPV_VALUE</td>
<td>VARCHAR2 (500)</td>
<td>Object attribute value</td>
</tr>
<tr>
<td>ERL_POLICY_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Entity extraction policy name</td>
</tr>
<tr>
<td>ERL_RULE_ID</td>
<td>INTEGER</td>
<td>Entity extraction rule ID</td>
</tr>
<tr>
<td>ERL_LANGUAGE</td>
<td>VARCHAR2 (30)</td>
<td>Entity extraction rule language</td>
</tr>
<tr>
<td>ERL_RULE</td>
<td>VARCHAR2 (512)</td>
<td>Entity extraction rule contents</td>
</tr>
<tr>
<td>ERL_TYPE</td>
<td>VARCHAR2 (4000)</td>
<td>String mapping backreferences to entity types</td>
</tr>
<tr>
<td>ERL_STATUS</td>
<td>VARCHAR2 (30)</td>
<td>Entity extraction rule status: compiled, not compiled, to be deleted.</td>
</tr>
<tr>
<td>ERL_COMMENTS</td>
<td>VARCHAR2 (4000)</td>
<td>Comments</td>
</tr>
<tr>
<td>ESE_POLICY_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Entity extraction policy name</td>
</tr>
<tr>
<td>ESE_NAME</td>
<td>VARCHAR2 (512)</td>
<td>Stop entity name</td>
</tr>
<tr>
<td>ESE_TYPE</td>
<td>VARCHAR2 (30)</td>
<td>Stop entity type</td>
</tr>
<tr>
<td>ESE_STATUS</td>
<td>VARCHAR2 (30)</td>
<td>Entity extraction rule status: compiled, not compiled, to be deleted, subset</td>
</tr>
<tr>
<td>ESE_COMMENTS</td>
<td>VARCHAR2 (4000)</td>
<td>Comments</td>
</tr>
</tbody>
</table>
**CTX_USER_FILTER_BY_COLUMNS**

This view displays all FILTER BY columns registered in the Text data dictionary for the current user. It can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBC_INDEX_NAME</td>
<td>VARCHAR2(30)</td>
<td>Index name</td>
</tr>
<tr>
<td>FBC_TABLE_OWNER</td>
<td>VARCHAR2(30)</td>
<td>Table owner name</td>
</tr>
<tr>
<td>FBC_TABLE_NAME</td>
<td>VARCHAR2(30)</td>
<td>Table name</td>
</tr>
<tr>
<td>FBC_COLUMN_NAME</td>
<td>VARCHAR2(256)</td>
<td>Column name</td>
</tr>
<tr>
<td>FBC_COLUMN_TYPE</td>
<td>VARCHAR2(30)</td>
<td>Column type</td>
</tr>
<tr>
<td>FBC_SECTION_TYPE</td>
<td>VARCHAR2(30)</td>
<td>Section type</td>
</tr>
<tr>
<td>FBC_SECTION_NAME</td>
<td>VARCHAR2(30)</td>
<td>Section name</td>
</tr>
<tr>
<td>FBC_SECTION_ID</td>
<td>NUMBER</td>
<td>Section ID</td>
</tr>
</tbody>
</table>

**CTX_USER_INDEXES**

This view displays all indexes that are registered in the Text data dictionary for the current user. It can be queried by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDX_CHARSET_COLUMN</td>
<td>VARCHAR2(256)</td>
<td>Name of the charset column of base table.</td>
</tr>
<tr>
<td>IDX_DOCID_COUNT</td>
<td>NUMBER</td>
<td>Number of documents indexed.</td>
</tr>
<tr>
<td>IDX_FORMAT_COLUMN</td>
<td>VARCHAR2(256)</td>
<td>Name of the format column of base table.</td>
</tr>
<tr>
<td>IDX_ID</td>
<td>NUMBER</td>
<td>Internal index ID.</td>
</tr>
<tr>
<td>IDX_KEY_NAME</td>
<td>VARCHAR(256)</td>
<td>Primary key column(s).</td>
</tr>
<tr>
<td>IDX_LANGUAGE_COLUMN</td>
<td>VARCHAR2(256)</td>
<td>Name of the language column of base table.</td>
</tr>
<tr>
<td>IDX_NAME</td>
<td>VARCHAR2(30)</td>
<td>Name of index.</td>
</tr>
<tr>
<td>IDX_STATUS</td>
<td>VARCHAR2(12)</td>
<td>Status, either INDEXED or INDEXING.</td>
</tr>
<tr>
<td>IDX_SYNC_INTERVAL</td>
<td>VARCHAR2(2000)</td>
<td>This is the interval string required by scheduler job. Only meaningful for AUTOMATIC sync. Always null for MANUAL and ON COMMIT sync.</td>
</tr>
<tr>
<td>IDX_SYNC_JOBNAME</td>
<td>VARCHAR2(50)</td>
<td>This is the scheduler job name for automatic sync. Only meaningful for AUTOMATIC sync and always null for other types of sync.</td>
</tr>
<tr>
<td>IDX_SYNC_MEMORY</td>
<td>VARCHAR2(100)</td>
<td>The sync memory size. Only meaningful for ON COMMIT and AUTOMATIC types of sync. For MANUAL sync, this is always null.</td>
</tr>
</tbody>
</table>
### CTX_USER_INDEX_PARTITIONS

This view displays all index partitions for the current user. It is queryable by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXP_DOCID_COUNT</td>
<td>NUMBER (38)</td>
<td>Number of documents associated with the index partition.</td>
</tr>
<tr>
<td>IXP_ID</td>
<td>NUMBER (38)</td>
<td>Index partition ID.</td>
</tr>
<tr>
<td>IXP_INDEX_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Index name.</td>
</tr>
</tbody>
</table>

### CTX_USER_INDEX_ERRORS

This view displays the indexing errors for the current user and is queryable by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERR_INDEX_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Name of index.</td>
</tr>
<tr>
<td>ERR_TIMESTAMP</td>
<td>DATE</td>
<td>Time of error.</td>
</tr>
<tr>
<td>ERR_TEXTKEY</td>
<td>VARCHAR2 (18) ROWID</td>
<td>ROWID of errored document or name of errored operation (for example, ALTER INDEX)</td>
</tr>
<tr>
<td>ERR_TEXT</td>
<td>VARCHAR2 (4000)</td>
<td>Error text.</td>
</tr>
</tbody>
</table>

### CTX_USER_INDEX_OBJECTS

This view displays the preferences that are attached to the indexes defined for the current user. It can be queried by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I XO_INDEX_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Name of index</td>
</tr>
<tr>
<td>I XO_CLASS</td>
<td>VARCHAR2 (30)</td>
<td>Object name</td>
</tr>
<tr>
<td>I XO_OBJECT</td>
<td>VARCHAR2 (80)</td>
<td>Object description</td>
</tr>
</tbody>
</table>

### Column Definitions

- **IDX_SYNC_PARA_DEGREE** (NUMBER): Degree of parallelism for sync. Only meaningful for the AUTOMATIC type of sync; always null for MANUAL and ON COMMIT syncs.
- **IDX_SYNC_TYPE** (VARCHAR2 (20)): Type of synching: AUTOMATIC, MANUAL or ON COMMIT.
- **IDX_TABLE** (VARCHAR2 (30)): Table name.
- **IDX_TABLE_OWNER** (VARCHAR2 (30)): Owner of table.
- **IDX_TEXT_NAME** (VARCHAR2 (30)): Text column name.
- **IDX_TYPE** (VARCHAR2 (30)): Type of index: CONTEXT, CTXCAT, or CTXRULE.
### CTX_USER_INDEX_SETS

This view displays all index set names that belong to the current user. It is queryable by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXS_NAME</td>
<td>VARCHAR2(30)</td>
<td>Index set name.</td>
</tr>
</tbody>
</table>

### CTX_USER_INDEX_SET_INDEXES

This view displays all the indexes in an index set that belong to the current user. It is queryable by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXX_INDEX_SET_NAME</td>
<td>VARCHAR2(30)</td>
<td>Index set name.</td>
</tr>
<tr>
<td>IXX_COLLIST</td>
<td>VARCHAR2(500)</td>
<td>Column list of the index.</td>
</tr>
<tr>
<td>IXX_STORAGE</td>
<td>VARCHAR2(500)</td>
<td>Storage clause of the index.</td>
</tr>
</tbody>
</table>

### CTX_USER_INDEX_SUB_LEXERS

This view shows the sub-lexers for each language for each index for the querying user. This view can be queried by all users.
**CTX_USER_INDEX_SUB_LEXER_VALS**

Shows the sub-lexer attributes and their values for the querying user. This view can be queried by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISL_INDEX_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Index name.</td>
</tr>
<tr>
<td>ISL_LANGUAGE</td>
<td>VARCHAR2 (30)</td>
<td>Language of sub-lexer</td>
</tr>
<tr>
<td>ISL_ALT_VALUE</td>
<td>VARCHAR2 (30)</td>
<td>Alternate value of language.</td>
</tr>
<tr>
<td>ISL_OBJECT</td>
<td>VARCHAR2 (30)</td>
<td>Name of lexer object used for this language.</td>
</tr>
<tr>
<td>ISV_INDEX_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Index name.</td>
</tr>
<tr>
<td>ISV_LANGUAGE</td>
<td>VARCHAR2 (30)</td>
<td>Language of sub-lexer</td>
</tr>
<tr>
<td>ISV_OBJECT</td>
<td>VARCHAR2 (30)</td>
<td>Name of lexer object used for this language.</td>
</tr>
<tr>
<td>ISV_ATTRIBUTE</td>
<td>VARCHAR2 (30)</td>
<td>Name of sub-lexer attribute.</td>
</tr>
<tr>
<td>ISV_VALUE</td>
<td>VARCHAR2 (500)</td>
<td>Value of sub-lexer attribute.</td>
</tr>
</tbody>
</table>

**CTX_USER_INDEX_VALUES**

This view displays attribute values for each object used in indexes for the current user. This view is queryable by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IXV_INDEX_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Index name.</td>
</tr>
<tr>
<td>IXV_CLASS</td>
<td>VARCHAR2 (30)</td>
<td>Class name.</td>
</tr>
<tr>
<td>IXV_OBJECT</td>
<td>VARCHAR2 (30)</td>
<td>Object name.</td>
</tr>
<tr>
<td>IXV_ATTRIBUTE</td>
<td>VARCHAR2 (30)</td>
<td>Attribute name</td>
</tr>
<tr>
<td>IXV_VALUE</td>
<td>VARCHAR2 (500)</td>
<td>Attribute value</td>
</tr>
</tbody>
</table>

**CTX_USER_ORDER_BY_COLUMNS**

This view displays all ORDER BY columns registered in the Text data dictionary for the current user. It can be queried by any user.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBC_INDEX_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Index name.</td>
</tr>
<tr>
<td>OBC_TABLE_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Table owner</td>
</tr>
<tr>
<td>OBC_TABLE_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Table name</td>
</tr>
<tr>
<td>OBC_COLUMN_NAME</td>
<td>VARCHAR2 (236)</td>
<td>Column name</td>
</tr>
<tr>
<td>OBC_COLUMN_POSITION</td>
<td>VARCHAR2 (30)</td>
<td>Column position</td>
</tr>
<tr>
<td>OBC_COLUMN_TYPE</td>
<td>VARCHAR2 (30)</td>
<td>Column type</td>
</tr>
</tbody>
</table>
This view displays a row for each of the user’s entries in the DML Queue. It can be queried by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBC_SECTION_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Section name</td>
</tr>
<tr>
<td>OBC_SECTION_TYPE</td>
<td>VARCHAR2 (30)</td>
<td>Section type</td>
</tr>
<tr>
<td>OBC_SECTION_ID</td>
<td>NUMBER</td>
<td>Section ID</td>
</tr>
<tr>
<td>OBC_SORT_ORDER</td>
<td>VARCHAR2 (8)</td>
<td>Sort order</td>
</tr>
</tbody>
</table>

**CTX_USER PREFERENCES**

This view displays all preferences defined by the current user. It can be queried by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Preference name.</td>
</tr>
<tr>
<td>PRE_CLASS</td>
<td>VARCHAR2 (30)</td>
<td>Preference class.</td>
</tr>
<tr>
<td>PRE_OBJECT</td>
<td>VARCHAR2 (30)</td>
<td>Object used.</td>
</tr>
</tbody>
</table>

**CTX_USER PREFERENCE_VALUES**

This view displays all the values for preferences defined by the current user. It can be queried by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRV_PREFERENCE</td>
<td>VARCHAR2 (30)</td>
<td>Preference name.</td>
</tr>
<tr>
<td>PRV_ATTRIBUTE</td>
<td>VARCHAR2 (64)</td>
<td>Attribute name</td>
</tr>
<tr>
<td>PRV_VALUE</td>
<td>VARCHAR2 (500)</td>
<td>Attribute value</td>
</tr>
</tbody>
</table>
**CTX_USERSECTION_GROUPS**

This view displays information about the section groups that have been created in the Text data dictionary for the current user. It can be queried by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGP_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Name of section group</td>
</tr>
<tr>
<td>SGP_TYPE</td>
<td>VARCHAR2 (30)</td>
<td>Type of section group</td>
</tr>
</tbody>
</table>

**CTX_USER_SQES**

This view displays the definitions of all persistent duration SQEs that have been created by the current user. In other words, it does not display session duration SQEs.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQE_OWNER</td>
<td>VARCHAR2 (30)</td>
<td>Owner of SQE</td>
</tr>
<tr>
<td>SQE_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Name of SQE</td>
</tr>
<tr>
<td>SQE_QUERY</td>
<td>CLOB</td>
<td>Query text</td>
</tr>
</tbody>
</table>

**CTX_USER_STOPLISTS**

This view displays stoplists for current user. It is queryable by all users.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPL_NAME</td>
<td>VARCHAR2 (30)</td>
<td>Name of stoplist.</td>
</tr>
<tr>
<td>SPL_COUNT</td>
<td>NUMBER</td>
<td>Number of stopwords</td>
</tr>
<tr>
<td>SPL_TYPE</td>
<td>VARCHAR2 (30)</td>
<td>Type of stoplist, MULTI or BASIC.</td>
</tr>
</tbody>
</table>

**CTX_USER_STOPWORDS**

This view displays stopwords in each stoplist for current user. Queryable by all users.
For the current user, this view contains information on multi-lexers and the sub-lexer preferences they contain. It can be queried by any user.

### Column Name | Type | Description
--- | --- | ---
SPW_STOPLIST | VARCHAR2 (30) | Stoplist name.
SPW_TYPE | VARCHAR2 (10) | Stop type, either STOP_WORD, STOP_CLASS, STOP_THEME.
SPW_WORD | VARCHAR2 (80) | Stopword.
SPW_LANGUAGE | VARCHAR2 (30) | Stopword language.

For the current user, this view displays the information about all of the thesauri that have been created in the system by the current user. It can be viewed by all users.

### Column Name | Type | Description
--- | --- | ---
THS_NAME | VARCHAR2 (30) | Thesaurus name

This view displays the phrase information of all thesauri owned by the current user. It can be queried by all users.

### Column Name | Type | Description
--- | --- | ---
THP_THESAURUS | VARCHAR2 (30) | Thesaurus name.
THP_PHRASE | VARCHAR2 (256) | Thesaurus phrase.
THP_QUALIFIER | VARCHAR2 (256) | Phrase qualifier.

This view displays the CTXSYS data dictionary and code version number information.

### Column Name | Type | Description
--- | --- | ---
VER_DICT | CHAR (9) | The CTXSYS data dictionary version number.
<table>
<thead>
<tr>
<th>Column Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VER_CODE</td>
<td>VARCHAR2(9)</td>
<td>The version number of the code linked in to the Oracle Database shadow process. This column fetches the version number for linked-in code. Thus, use this column to detect and verify patch releases.</td>
</tr>
</tbody>
</table>
Stopword Transformations in Oracle Text

This appendix describes stopword transformations. The following topic is covered:

- Understanding Stopword Transformations

Understanding Stopword Transformations

When you use a stopword or stopword-only phrase as an operand for a query operator, Oracle Text rewrites the expression to eliminate the stopword or stopword-only phrase and then executes the query.

The following section describes the stopword rewrites or transformations for each operator. In all tables, the Stopword Expression column describes the query expression or component of a query expression, while the right-hand column describes the way Oracle Text rewrites the query.

The token stopword stands for a single stopword or a stopword-only phrase.

The token non_stopword stands for either a single non-stopword, a phrase of all non-stopwords, or a phrase of non-stopwords and stopwords.

The token no_lex stands for a single character or a string of characters that is neither a stopword nor a word that is indexed. For example, the + character by itself is an example of a no_lex token.

When the Stopword Expression column completely describes the query expression, a rewritten expression of no_token means that no hits are returned when you enter such a query.

When the Stopword Expression column describes a component of a query expression with more than one operator, a rewritten expression of no_token means that a no_token value is passed to the next step of the rewrite.

Transformations that contain a no_token as an operand in the Stopword Expression column describe intermediate transformations in which the no_token is a result of a previous transformation. These intermediate transformations apply when the original query expression has at least one stopword and more than one operator.

For example, consider the following compound query expression:

'\{this NOT dog\} AND cat'

Assuming that this is the only stopword in this expression, Oracle Text applies the following transformations in the following order:

stopword NOT non-stopword => no_token

no_token AND non_stopword => non_stopword
The resulting expression is:

'cat'

**Word Transformations**

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>stopword</td>
<td>no_token</td>
</tr>
<tr>
<td>no_lex</td>
<td>no_token</td>
</tr>
</tbody>
</table>

The first transformation means that a stopword or stopword-only phrase by itself in a query expression results in no hits.

The second transformation says that a term that is not lexed, such as the + character, results in no hits.

**AND Transformations**

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>non_stopword AND stopword</td>
<td>non_stopword</td>
</tr>
<tr>
<td>non_stopword AND no_token</td>
<td>non_stopword</td>
</tr>
<tr>
<td>stopword AND non_stopword</td>
<td>non_stopword</td>
</tr>
<tr>
<td>no_token AND non_stopword</td>
<td>non_stopword</td>
</tr>
<tr>
<td>stopword AND stopword</td>
<td>no_token</td>
</tr>
<tr>
<td>no_token AND stopword</td>
<td>no_token</td>
</tr>
<tr>
<td>stopword AND no_token</td>
<td>no_token</td>
</tr>
<tr>
<td>no_token AND no_token</td>
<td>no_token</td>
</tr>
</tbody>
</table>

**OR Transformations**

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>non_stopword OR stopword</td>
<td>non_stopword</td>
</tr>
<tr>
<td>non_stopword OR no_token</td>
<td>non_stopword</td>
</tr>
<tr>
<td>stopword OR non_stopword</td>
<td>non_stopword</td>
</tr>
<tr>
<td>no_token OR non_stopword</td>
<td>non_stopword</td>
</tr>
<tr>
<td>stopword OR stopword</td>
<td>no_token</td>
</tr>
<tr>
<td>no_token OR stopword</td>
<td>no_token</td>
</tr>
<tr>
<td>stopword OR no_token</td>
<td>no_token</td>
</tr>
<tr>
<td>no_token OR no_token</td>
<td>no_token</td>
</tr>
</tbody>
</table>
### ACCUMulate Transformations

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>non_stopword ACCUM stopword</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>non_stopword ACCUM no_token</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>stopword ACCUM non_stopword</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>no_token ACCUM non_stopword</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>stopword ACCUM stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token ACCUM stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>stopword ACCUM no_token</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token ACCUM no_token</code></td>
<td><code>no_token</code></td>
</tr>
</tbody>
</table>

### MINUS Transformations

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>non_stopword MINUS stopword</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>non_stopword MINUS no_token</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>stopword MINUS non_stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token MINUS non_stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>stopword MINUS stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token MINUS stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>stopword MINUS no_token</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token MINUS no_token</code></td>
<td><code>no_token</code></td>
</tr>
</tbody>
</table>

### MNOT Transformations

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>non_stopword MNOT stopword</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>non_stopword MNOT no_token</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>stopword MNOT non_stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token MNOT non_stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>stopword MNOT stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token MNOT stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>stopword MNOT no_token</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token MNOT no_token</code></td>
<td><code>no_token</code></td>
</tr>
</tbody>
</table>

### NOT Transformations

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>non_stopword NOT stopword</code></td>
<td><code>non_stopword</code></td>
</tr>
</tbody>
</table>
### Understanding Stopword Transformations

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>non_stopword NOT no_token</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>stopword NOT non_stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token NOT non_stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>stopword NOT stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token NOT stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>stopword NOT no_token</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token NOT no_token</code></td>
<td><code>no_token</code></td>
</tr>
</tbody>
</table>

### EQUIValence Transformations

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>non_stopword EQUIV stopword</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>non_stopword EQUIV no_token</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>stopword EQUIV non_stopword</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>no_token EQUIV non_stopword</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>stopword EQUIV stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token EQUIV stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>stopword EQUIV no_token</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token EQUIV no_token</code></td>
<td><code>no_token</code></td>
</tr>
</tbody>
</table>

**Note:** When you use query explain plan, not all of the equivalence transformations are represented in the `EXPLAIN` table.

### NEAR Transformations

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>non_stopword NEAR stopword</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>non_stopword NEAR no_token</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>stopword NEAR non_stopword</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>no_token NEAR non_stopword</code></td>
<td><code>non_stopword</code></td>
</tr>
<tr>
<td><code>stopword NEAR stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token NEAR stopword</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>stopword NEAR no_token</code></td>
<td><code>no_token</code></td>
</tr>
<tr>
<td><code>no_token NEAR no_token</code></td>
<td><code>no_token</code></td>
</tr>
</tbody>
</table>
### Weight Transformations

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>stopword * n</code></td>
<td>no_token</td>
</tr>
<tr>
<td><code>no_token * n</code></td>
<td>no_token</td>
</tr>
</tbody>
</table>

### Threshold Transformations

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>stopword &gt; n</code></td>
<td>no_token</td>
</tr>
<tr>
<td><code>no_token &gt; n</code></td>
<td>no_token</td>
</tr>
</tbody>
</table>

### WITHIN Transformations

<table>
<thead>
<tr>
<th>Stopword Expression</th>
<th>Rewritten Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>stopword WITHIN section</code></td>
<td>no_token</td>
</tr>
<tr>
<td><code>no_token WITHIN section</code></td>
<td>no_token</td>
</tr>
</tbody>
</table>
This appendix provides the parts-of-speech tagging for languages used with AUTO_LEXER. This appendix contains quick references for the following languages:

- Tagging in Arabic
- Tagging in Catalan
- Tagging in Chinese - Traditional and Simplified
- Tagging in Croatian
- Tagging in Danish
- Tagging in Dutch
- Tagging in English
- Tagging in Farsi
- Tagging in Finnish
- Tagging in French
- Tagging in German
- Tagging in Italian
- Tagging in Japanese
- Tagging in Korean
- Tagging in Bokmal
- Tagging in Nynorsk
- Tagging in Portuguese
- Tagging in Russian
- Tagging in Slovak
- Tagging in Slovenian
- Tagging in Spanish
- Tagging in Swedish

**Tagging in Arabic**

The charts in this section provide a reference for tagging in Arabic.
### Figure I–1 Tagging in Arabic (Abbr - Nn-Masc)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbr</td>
<td>Abbreviation</td>
<td>ان بي سي</td>
</tr>
<tr>
<td>Adj-ConjDet-Fem</td>
<td>Feminine adjective with attached conjunction and determiner</td>
<td>والسياسة</td>
</tr>
<tr>
<td>Adj-Det-Fem</td>
<td>Feminine adjective with attached determiner</td>
<td>الإدارة</td>
</tr>
<tr>
<td>Adj-ConjDet-Masc</td>
<td>Masculine adjective with attached conjunction and determiner</td>
<td>والاقتصاد</td>
</tr>
<tr>
<td>Adj-Det-Masc</td>
<td>Masculine adjective with attached determiner</td>
<td>الأول</td>
</tr>
<tr>
<td>Adj-Conj-Fem</td>
<td>Feminine adjective with attached conjunction</td>
<td>وسيلة</td>
</tr>
<tr>
<td>Adj-Fem</td>
<td>Feminine adjective</td>
<td>متغيرة</td>
</tr>
<tr>
<td>Adj-Conj-Masc</td>
<td>Masculine adjective with attached conjunction</td>
<td>وطني موليلي</td>
</tr>
<tr>
<td>Adj-Masc</td>
<td>Masculine adjective</td>
<td>أبرزها</td>
</tr>
<tr>
<td>Adj Öl</td>
<td>Adjective</td>
<td>واضح</td>
</tr>
<tr>
<td>Adv</td>
<td>Adverb</td>
<td>فم</td>
</tr>
<tr>
<td>Conj</td>
<td>Detached conjunction</td>
<td>عندما</td>
</tr>
<tr>
<td>Det</td>
<td>Detached determiner</td>
<td>لل</td>
</tr>
<tr>
<td>Fw</td>
<td>Function word</td>
<td>إن</td>
</tr>
<tr>
<td>Interj</td>
<td>Interjection</td>
<td>يا</td>
</tr>
<tr>
<td>Nn-ConjDet-Fem</td>
<td>Feminine noun with attached conjunction and determiner</td>
<td>المؤسسة</td>
</tr>
<tr>
<td>Nn-Det-Fem</td>
<td>Feminine noun with attached determiner</td>
<td>الصفحة</td>
</tr>
<tr>
<td>Nn-ConjDet-Masc</td>
<td>Masculine noun with attached conjunction and determiner</td>
<td>والمواطنين</td>
</tr>
<tr>
<td>Nn-Det-Masc</td>
<td>Masculine noun with attached determiner</td>
<td>الكهرباء</td>
</tr>
<tr>
<td>Nn-Conj-Fem</td>
<td>Feminine noun with attached conjunction</td>
<td>الانتخابات</td>
</tr>
<tr>
<td>Nn-Fem</td>
<td>Feminine noun</td>
<td>تهديدات</td>
</tr>
<tr>
<td>Nn-Conj-Masc</td>
<td>Masculine noun with attached conjunction</td>
<td>عملا</td>
</tr>
<tr>
<td>Nn-Masc</td>
<td>Masculine noun</td>
<td>مجلس</td>
</tr>
</tbody>
</table>
### Figure I–2  Tagging in Arabic (Nn-Phone - Punct)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nn-Phone</td>
<td>Noun indicating a phone number follows</td>
<td>نموذج</td>
</tr>
<tr>
<td>Num-Phone</td>
<td>Number</td>
<td>235</td>
</tr>
<tr>
<td>Num-Date</td>
<td>Numeric date</td>
<td>2005/5/23</td>
</tr>
<tr>
<td>Num-Decimal</td>
<td>Decimal number</td>
<td>100.0 143.5</td>
</tr>
<tr>
<td>Num-Time</td>
<td>Numeric time expression</td>
<td>10:45:00 23:58</td>
</tr>
<tr>
<td>Num-Gregorian-Year</td>
<td>Numeric Gregorian year</td>
<td>2003م</td>
</tr>
<tr>
<td>Num-Hijri-Year</td>
<td>Numeric Hijri year</td>
<td>1420ه</td>
</tr>
<tr>
<td>Num-Percent</td>
<td>Numeric percent expression</td>
<td>45.5%-200%</td>
</tr>
<tr>
<td>Part-Fut</td>
<td>Future particle</td>
<td>سوف</td>
</tr>
<tr>
<td>Part-Inf</td>
<td>Interrogative particle</td>
<td>هل</td>
</tr>
<tr>
<td>Part-Neg</td>
<td>Negation particle</td>
<td>لا</td>
</tr>
<tr>
<td>Prop</td>
<td>Proposition</td>
<td>في</td>
</tr>
<tr>
<td>Pron-1P</td>
<td>First person pronoun</td>
<td>نحن</td>
</tr>
<tr>
<td>Pron-2P-Fem</td>
<td>Second person feminine pronoun</td>
<td>لك</td>
</tr>
<tr>
<td>Pron-2P-Masc</td>
<td>Second person masculine pronoun</td>
<td>نحن</td>
</tr>
<tr>
<td>Pron-3P</td>
<td>Third person pronoun</td>
<td>هؤلاء</td>
</tr>
<tr>
<td>Pron-3P-Fem</td>
<td>Third person feminine pronoun</td>
<td>هؤلاء</td>
</tr>
<tr>
<td>Pron-3P-Masc</td>
<td>Third person masculine pronoun</td>
<td>هؤلاء</td>
</tr>
<tr>
<td>Pron-Dim-Fem</td>
<td>Demonstrative feminine pronoun</td>
<td>هؤلاء</td>
</tr>
<tr>
<td>Pron-Dim-Masc</td>
<td>Demonstrative masculine pronoun</td>
<td>هؤلاء</td>
</tr>
<tr>
<td>Pron-Rel</td>
<td>Relative pronoun</td>
<td>الين</td>
</tr>
<tr>
<td>Prop</td>
<td>Proper noun</td>
<td>زهير</td>
</tr>
<tr>
<td>Prop-Country</td>
<td>Proper noun that is a country name</td>
<td>مصر</td>
</tr>
<tr>
<td>Prop-Email</td>
<td>Email address</td>
<td><a href="mailto:info@taxight.com">info@taxight.com</a></td>
</tr>
<tr>
<td>Prop-URL</td>
<td>URL address</td>
<td><a href="mailto:johndoe@aboxyz.co.uk">johndoe@aboxyz.co.uk</a></td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>Comma</td>
<td>،</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>Sentence ending punctuation</td>
<td>و</td>
</tr>
<tr>
<td>Punct</td>
<td>Punctuation</td>
<td>،</td>
</tr>
</tbody>
</table>

### Figure I–3  Tagging in Arabic (V-Imperativ - V-Perf)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>V-Imperativ</td>
<td>Imperative Verb</td>
<td>ايع</td>
</tr>
<tr>
<td>V-Imperf</td>
<td>Imperfect verb</td>
<td>تستجب</td>
</tr>
<tr>
<td>V-Perf</td>
<td>Perfect verb</td>
<td>وقال</td>
</tr>
</tbody>
</table>

### Tagging in Catalan

The charts in this sections provide a reference for tagging in Catalan.
**Figure I–4  Tagging in Catalan (Abbr to Det-Dem-Sg)**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbr</td>
<td>Abbreviation</td>
<td>etc, Cia,</td>
</tr>
<tr>
<td>Adj</td>
<td>Adjective, invariant for number</td>
<td>alfín, antídrola</td>
</tr>
<tr>
<td>Adj-Ord-Pl</td>
<td>Plural, spelled out ordinal adjective</td>
<td>segons, primeres</td>
</tr>
<tr>
<td>Adj-Ord-Sg</td>
<td>Singular, spelled out ordinal adjective</td>
<td>primer, segon</td>
</tr>
<tr>
<td>Adj-Pl</td>
<td>Plural adjective</td>
<td>factibles</td>
</tr>
<tr>
<td>Adv-Sg</td>
<td>Singular adjective</td>
<td>absurd, calent, capoç</td>
</tr>
<tr>
<td>Adv</td>
<td>Adverb</td>
<td>aci, abans, gairebé, fins</td>
</tr>
<tr>
<td>Adv-Deg</td>
<td>Adverbs that can modify adjectives</td>
<td>bastant, força, gaire, massa, més, mig, molt, poc, prov, ton, tot</td>
</tr>
<tr>
<td>Adv-Int</td>
<td>Interrogative Adverbs</td>
<td>quan, on, a on, d'on, con, per què</td>
</tr>
<tr>
<td>Adv-Rel</td>
<td>Adverbial Relativizer</td>
<td>quan, con, on</td>
</tr>
<tr>
<td>Aux-Inf-be</td>
<td>Infinitive 'ser'</td>
<td>ser</td>
</tr>
<tr>
<td>Aux-Inf-have</td>
<td>Infinitive 'haver'</td>
<td>haver</td>
</tr>
<tr>
<td>Aux-anar</td>
<td>Auxiliary 'anar'</td>
<td>vaig</td>
</tr>
<tr>
<td>Aux-be</td>
<td>Auxiliary 'ser'</td>
<td>serà</td>
</tr>
<tr>
<td>Aux-have</td>
<td>Auxiliary 'haver'</td>
<td>ha, han</td>
</tr>
<tr>
<td>Conq</td>
<td>Conjunction</td>
<td>si, perquè, mentre</td>
</tr>
<tr>
<td>Con-Coord</td>
<td>Coordinating conjunction</td>
<td>i, o, ni</td>
</tr>
<tr>
<td>Con-ccom</td>
<td>Conjunction 'com'</td>
<td>con</td>
</tr>
<tr>
<td>Con-que</td>
<td>Conjunction 'que'</td>
<td>que</td>
</tr>
<tr>
<td>Det-Def-Pl</td>
<td>Plural definite determiner</td>
<td>les, els</td>
</tr>
<tr>
<td>Det-Def-Sg</td>
<td>Singular definite determiner</td>
<td>et, la, l'</td>
</tr>
<tr>
<td>Det-Dem-Pl</td>
<td>Plural demonstrative determiner</td>
<td>aquesíes</td>
</tr>
<tr>
<td>Det-Dem-Sg</td>
<td>Singular demonstrative determiner</td>
<td>aquest, això</td>
</tr>
</tbody>
</table>
Figure I–5  Tagging in Catalan (Det-Indef-Pl to Punct-Close)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Det-Indef-Pl</td>
<td>Plural indefinite determiner or pronoun</td>
<td>gaires, molts, uns</td>
</tr>
<tr>
<td>Det-Indef-Sg</td>
<td>Singular indefinite determiner or pronoun</td>
<td>bastant, gaire, quant, tant, molt, poc</td>
</tr>
<tr>
<td>Det-Int-Pl</td>
<td>Plural interrogative determiner</td>
<td>quins</td>
</tr>
<tr>
<td>Det-Int-Sg</td>
<td>Singular interrogative determiner</td>
<td>quin, quan</td>
</tr>
<tr>
<td>Det-Poss-Pl</td>
<td>Plural possessive determiner</td>
<td>nostres, seves, llurs</td>
</tr>
<tr>
<td>Det-Poss-Sg</td>
<td>Singular possessive determiner</td>
<td>teu, una, lliur</td>
</tr>
<tr>
<td>Det-Rel-Sg</td>
<td>Singular relative determiner</td>
<td>qui, quell</td>
</tr>
<tr>
<td>Det-Rel-Pl</td>
<td>Plural relative determiner</td>
<td>quals</td>
</tr>
<tr>
<td>Interj</td>
<td>Interjection or onomatopoeia</td>
<td>uii, ehh?</td>
</tr>
<tr>
<td>Nn</td>
<td>Noun, invariable for number</td>
<td>atles, allotres, focus</td>
</tr>
<tr>
<td>Mn-Net</td>
<td>URL or e-mail address</td>
<td><a href="http://www.iaight.com">www.iaight.com</a>, <a href="mailto:info@iaight.com">info@iaight.com</a></td>
</tr>
<tr>
<td>Mn-Pl</td>
<td>Plural noun</td>
<td>organitzacions, xarxes, cases, drets</td>
</tr>
<tr>
<td>Mn-Sg</td>
<td>Singular noun</td>
<td>manera, exemple</td>
</tr>
<tr>
<td>Num</td>
<td>Numeric expression, or cardinal number</td>
<td>2001, millions, dos</td>
</tr>
<tr>
<td>Num-Ord</td>
<td>Ordinal number</td>
<td>1r, 2a, 3t, 4t, 5e</td>
</tr>
<tr>
<td>Part-Neg</td>
<td>The negation particle 'no'</td>
<td>no</td>
</tr>
<tr>
<td>Prep</td>
<td>Preposition</td>
<td>amb, a causa de, darrera, en</td>
</tr>
<tr>
<td>Prep-Det-a</td>
<td>Combination 'a' and determiner</td>
<td>al, als</td>
</tr>
<tr>
<td>Prep-Det-de</td>
<td>Combination 'de' and determiner</td>
<td>del, des, des del</td>
</tr>
<tr>
<td>Prep-Det-per</td>
<td>Combination 'per' and determiner</td>
<td>pel, peis</td>
</tr>
<tr>
<td>Prep-a</td>
<td>Preposition a</td>
<td>a</td>
</tr>
<tr>
<td>Prep-da</td>
<td>Preposition do</td>
<td>do, d</td>
</tr>
<tr>
<td>Prep-per</td>
<td>Preposition 'per'</td>
<td>per</td>
</tr>
<tr>
<td>Pron</td>
<td>Pronoun</td>
<td>jo, tu, ell, axò</td>
</tr>
<tr>
<td>Pron-Adv</td>
<td>Adverbial pronoun</td>
<td>en, bi, ni, -en</td>
</tr>
<tr>
<td>Pron-Clitic</td>
<td>Clitic pronoun</td>
<td>s', ns', -hi</td>
</tr>
<tr>
<td>Pron-Dem</td>
<td>Demonstrative pronoun</td>
<td>aquesta</td>
</tr>
<tr>
<td>Pron-Indef</td>
<td>Indefinite pronouns</td>
<td>moltes</td>
</tr>
<tr>
<td>Pron-Int</td>
<td>Interrogative and exclamative pronoun</td>
<td>qui, què, quant, quantes</td>
</tr>
<tr>
<td>Pron-Obliq</td>
<td>Oblique pronoun</td>
<td>en, ho, all, am</td>
</tr>
<tr>
<td>Pron-Ord</td>
<td>Ordinal pronoun</td>
<td>tercer</td>
</tr>
<tr>
<td>Pron-Poss</td>
<td>Possessive pronoun</td>
<td>el meu, la seva</td>
</tr>
<tr>
<td>Pron-Rel</td>
<td>Relative pronoun</td>
<td>qui, que, qui, qual</td>
</tr>
<tr>
<td>Pron-poss</td>
<td>'le' pronoun</td>
<td>es, se, s', s'</td>
</tr>
<tr>
<td>Prop</td>
<td>Proper noun or alpha numeric combination</td>
<td>Europa, FAO/OMS</td>
</tr>
<tr>
<td>Punct</td>
<td>Other punctuation</td>
<td>; ; ; ; ( &amp; /</td>
</tr>
<tr>
<td>Punct-Close</td>
<td>Closed parentheses</td>
<td>)</td>
</tr>
</tbody>
</table>
Tagging in Chinese - Traditional and Simplified

The chart in this section provide a reference for tagging in Traditional and Simplified Chinese.

### Figure I–6 Tagging in Catalan (Punct-Comma to V/Adj-PaPart-Sg)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punct-Comma</td>
<td>Comma</td>
<td></td>
</tr>
<tr>
<td>Punct-Open</td>
<td>open parenthesis</td>
<td>(</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>Sentence ending punctuation</td>
<td>... ! ?</td>
</tr>
<tr>
<td>V-Fin</td>
<td>Finite verb</td>
<td>recemen, recem, passa, va</td>
</tr>
<tr>
<td>V-Impv</td>
<td>Imperative verb</td>
<td>obtenim, traça</td>
</tr>
<tr>
<td>V-Inf</td>
<td>Infinitive verb</td>
<td>arribar, mantenir, buscar</td>
</tr>
<tr>
<td>V-PrPart</td>
<td>Present participle verb</td>
<td>present, essent, donant</td>
</tr>
<tr>
<td>V/Adj-PaPart-Pl</td>
<td>Plural post participle verb or adjective</td>
<td>ells, es, enclaves</td>
</tr>
<tr>
<td>V/Adj-PaPart-Sg</td>
<td>Singular post participle verb or adjective</td>
<td>fet, assenyalat, mancada</td>
</tr>
</tbody>
</table>

---

Tagging in Croatian

The charts in this section provide a reference for tagging in Croatian.
Tagging in Danish

The charts in this section provide a reference for tagging in Danish.
### Figure I–9  Tagging in Danish (Abbr to Conj-hvis)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbr</td>
<td>abbreviation</td>
<td>nov., kg., dkr., USA</td>
</tr>
<tr>
<td>Adj</td>
<td>adjective</td>
<td>det gode brød</td>
</tr>
<tr>
<td>Adj-Comp</td>
<td>comparative adjective</td>
<td>et bedre forslag</td>
</tr>
<tr>
<td>Adj-Gen</td>
<td>genitive adjective</td>
<td>den enkeltes tryghed</td>
</tr>
<tr>
<td>Adj-PaPart</td>
<td>past participle used as adjective</td>
<td>bestemt niveau</td>
</tr>
<tr>
<td>Adj-PaPart-Gen</td>
<td>post participle used as adjective, genitive</td>
<td>den ansat tes papirer</td>
</tr>
<tr>
<td>Adj-Part</td>
<td>present participle adjective</td>
<td>manglede</td>
</tr>
<tr>
<td>Adj-Sup</td>
<td>superlative adjective</td>
<td>fleste</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb (includes pronouns)</td>
<td>igen</td>
</tr>
<tr>
<td>Adv-Comp</td>
<td>comparative adverb</td>
<td>tidligere</td>
</tr>
<tr>
<td>Adv-IntRel</td>
<td>interrogative or relative adverb</td>
<td>hvor, hvorefter</td>
</tr>
<tr>
<td>Adv-Sup</td>
<td>superlative adverb</td>
<td>oftest</td>
</tr>
<tr>
<td>AuxV-Infin-blive</td>
<td>infinitive auxiliary or main verb blive</td>
<td>blive</td>
</tr>
<tr>
<td>AuxV-Infin-faa</td>
<td>infinitive auxiliary or main verb fæ</td>
<td>fæ</td>
</tr>
<tr>
<td>AuxV-Infin-have</td>
<td>infinitive auxiliary or main verb have</td>
<td>have</td>
</tr>
<tr>
<td>AuxV-Infin-være</td>
<td>infinitive auxiliary or main verb vare</td>
<td>vare</td>
</tr>
<tr>
<td>AuxV-PaPart-blive</td>
<td>past participle auxiliary or main verb blive</td>
<td>blivet</td>
</tr>
<tr>
<td>AuxV-PaPart-faa</td>
<td>past participle auxiliary or main verb fæ</td>
<td>fæet</td>
</tr>
<tr>
<td>AuxV-PaPart-have</td>
<td>past participle auxiliary or main verb have</td>
<td>havet</td>
</tr>
<tr>
<td>AuxV-PaPart-være</td>
<td>past participle auxiliary or main verb vare</td>
<td>varet</td>
</tr>
<tr>
<td>AuxV-Past-blive</td>
<td>past tense auxiliary or main verb blive</td>
<td>blev</td>
</tr>
<tr>
<td>AuxV-Past-faa</td>
<td>past tense auxiliary or main verb fæ</td>
<td>fik</td>
</tr>
<tr>
<td>AuxV-Past-have</td>
<td>past tense auxiliary or main verb have</td>
<td>havde</td>
</tr>
<tr>
<td>AuxV-Past-være</td>
<td>past tense auxiliary or main verb vare</td>
<td>vare</td>
</tr>
<tr>
<td>AuxV-Præ-blive</td>
<td>present tense auxiliary or main verb blive</td>
<td>bliver</td>
</tr>
<tr>
<td>AuxV-Præ-faa</td>
<td>present tense auxiliary or main verb fæ</td>
<td>fær</td>
</tr>
<tr>
<td>AuxV-Præ-have</td>
<td>present tense auxiliary or main verb have</td>
<td>hær</td>
</tr>
<tr>
<td>AuxV-Præ-være</td>
<td>present tense auxiliary or main verb vare</td>
<td>er</td>
</tr>
<tr>
<td>Comp-Part</td>
<td>left compound part</td>
<td>post- og telegrafvæsenet</td>
</tr>
<tr>
<td>Coni</td>
<td>conjunction</td>
<td>at, når</td>
</tr>
<tr>
<td>Coni-Coord</td>
<td>coordinating conjunction</td>
<td>og, eller</td>
</tr>
<tr>
<td>Coni-hvis</td>
<td>conjunction or relative pronoun hvis</td>
<td>hvis</td>
</tr>
</tbody>
</table>
### Figure I–10  Tagging in Danish (Conj-som to Punct-Comma)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conj-som</td>
<td>conjunction or relative pronoun som</td>
<td>som</td>
</tr>
<tr>
<td>Det</td>
<td>determiner</td>
<td>en</td>
</tr>
<tr>
<td>Det-Indet</td>
<td>indeterminate determiner</td>
<td>forskellig, somme</td>
</tr>
<tr>
<td>Det-Indet-Gen</td>
<td>indeterminate determiner, genitive</td>
<td>forskelliges</td>
</tr>
<tr>
<td>Det-Coord</td>
<td>conjunctional adverb</td>
<td>både</td>
</tr>
<tr>
<td>Det/Pron-Int/Rel</td>
<td>interrogative or relative pronoun</td>
<td>hvad, hvem, hvilke</td>
</tr>
<tr>
<td>Det/Pron-Poss</td>
<td>possessive determiner or pronoun</td>
<td>vores, min</td>
</tr>
<tr>
<td>Det/Pron-Poss-Ref</td>
<td>reflexive possessive pronoun</td>
<td>sin, sit, sine</td>
</tr>
<tr>
<td>Det/Pron-Quant</td>
<td>quantifying determiner or pronoun</td>
<td>mange</td>
</tr>
<tr>
<td>Det/Pron-Quant-Com</td>
<td>comparative more</td>
<td>mere</td>
</tr>
<tr>
<td>Det/Pron-Quant-Gen</td>
<td>genitive quantifying determiner or pronoun</td>
<td>mange, alle</td>
</tr>
<tr>
<td>Det/Pron-Quant-Pre</td>
<td>quantifying pre-determiner or pronoun</td>
<td>hver, hver</td>
</tr>
<tr>
<td>Det/Pron-Quant-Sup</td>
<td>superlative most</td>
<td>mest</td>
</tr>
<tr>
<td>Func</td>
<td>function word (miscellaneous category)</td>
<td>ambulanci, børne</td>
</tr>
<tr>
<td>Interj</td>
<td>interjection</td>
<td>hj</td>
</tr>
<tr>
<td>Nn</td>
<td>noun</td>
<td>kvinde</td>
</tr>
<tr>
<td>Nn-Gen</td>
<td>genitive noun</td>
<td>kvindens</td>
</tr>
<tr>
<td>Nn-Letter</td>
<td>lowercase and uppercase letters</td>
<td>b, N</td>
</tr>
<tr>
<td>Nn-Net</td>
<td>URL and e-mail address</td>
<td><a href="http://www.insight.com">www.insight.com</a></td>
</tr>
<tr>
<td>Num</td>
<td>cardinal number (in digits or words)</td>
<td>3, tre</td>
</tr>
<tr>
<td>Ord</td>
<td>ordinal number, in digits or spelled out</td>
<td>20., femtøn</td>
</tr>
<tr>
<td>Part-inf</td>
<td>inflexional particle at</td>
<td>få lov til attte</td>
</tr>
<tr>
<td>Part-Neg</td>
<td>negative particle</td>
<td>ikke</td>
</tr>
<tr>
<td>Prep</td>
<td>preposition</td>
<td>med, hos</td>
</tr>
<tr>
<td>Prep-af</td>
<td>preposition after</td>
<td>af</td>
</tr>
<tr>
<td>Pron</td>
<td>pronoun</td>
<td>den, denne</td>
</tr>
<tr>
<td>Pron-Expl</td>
<td>expressive pronoun</td>
<td>drikke</td>
</tr>
<tr>
<td>Pron-Gen</td>
<td>genitive pronoun</td>
<td>begge</td>
</tr>
<tr>
<td>Pron-Pers</td>
<td>personal pronoun</td>
<td>jeg, mig</td>
</tr>
<tr>
<td>Pron-Recip</td>
<td>reciprocal pronoun</td>
<td>hirunden</td>
</tr>
<tr>
<td>Pron-Recip-Gen</td>
<td>genitive reciprocal pronoun</td>
<td>hirundene</td>
</tr>
<tr>
<td>Pron-Pol</td>
<td>relative pronouns der and same</td>
<td>familier, der skulkes</td>
</tr>
<tr>
<td>Prop</td>
<td>proper name, initials or title</td>
<td>Ole, H., fra, dr.</td>
</tr>
<tr>
<td>Prop-Gen</td>
<td>genitive proper name</td>
<td>Jensens bil</td>
</tr>
<tr>
<td>Punct</td>
<td>miscellaneous punctuation</td>
<td>-</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>comma</td>
<td>,</td>
</tr>
</tbody>
</table>
Tagging in Dutch

The charts in this section provide a reference for tagging in Dutch.

Figure I–11  Tagging in Danish (Punct-Sent to V-Pres-SForm)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punct-Sent</td>
<td>sentence boundary punctuation</td>
<td>. ? !</td>
</tr>
<tr>
<td>V-Imprv</td>
<td>imperative verb</td>
<td>skriv</td>
</tr>
<tr>
<td>V-Infin</td>
<td>infinitive verb</td>
<td>skrive</td>
</tr>
<tr>
<td>V-PoPart</td>
<td>past participle verb</td>
<td>skrevet</td>
</tr>
<tr>
<td>V-Past</td>
<td>past tense verb</td>
<td>skrov</td>
</tr>
<tr>
<td>V-PastSForm</td>
<td>past tense S-form verb</td>
<td>tales</td>
</tr>
<tr>
<td>V-Pres</td>
<td>present tense verb</td>
<td>sker</td>
</tr>
<tr>
<td>V-PresSForm</td>
<td>present tense S-form verb</td>
<td>sendes</td>
</tr>
</tbody>
</table>

Figure I–12  Tagging in Dutch (Adj-Attr to Conj-Sub)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj-Attr</td>
<td>attributive adjective</td>
<td>een snelle auto</td>
</tr>
<tr>
<td>Adj-Post</td>
<td>postmodifying adjective</td>
<td>wat anders</td>
</tr>
<tr>
<td>Adj-Pred</td>
<td>predicative adjective</td>
<td>hij is snel</td>
</tr>
<tr>
<td>Adv</td>
<td>non-adjectival adverb</td>
<td>stroom opwaarts</td>
</tr>
<tr>
<td>Adv-Deg</td>
<td>adverbs that can modify adjectives</td>
<td>hij wil ze snel</td>
</tr>
<tr>
<td>Adv-Int</td>
<td>interrogative adverb</td>
<td>waarom gaat hij</td>
</tr>
<tr>
<td>Adv-Pron</td>
<td>pronominal adverb</td>
<td>hij praat hierover</td>
</tr>
<tr>
<td>Aux-Fin</td>
<td>finite auxiliary verb</td>
<td>hij is geweest</td>
</tr>
<tr>
<td>Aux-Inf</td>
<td>infinitive auxiliary verb</td>
<td>hij zal zijn</td>
</tr>
<tr>
<td>Aux-PaPart</td>
<td>past participle auxiliary verb</td>
<td>hij is geweest</td>
</tr>
<tr>
<td>Conpd-Left</td>
<td>left truncated part of compound</td>
<td>honden- en kattenvoer</td>
</tr>
<tr>
<td>Conpd-Right</td>
<td>right truncated part of compound</td>
<td>kattenvoer en -melk</td>
</tr>
<tr>
<td>Conj-Comp</td>
<td>comparative conjunction</td>
<td>zo groot als</td>
</tr>
<tr>
<td>Conj-Coord</td>
<td>coordinating conjunction</td>
<td>jan en marije</td>
</tr>
<tr>
<td>Conj-Inf</td>
<td>infinitive conjunction</td>
<td>door te vragen</td>
</tr>
<tr>
<td>Conj-Rel</td>
<td>relative conjunction</td>
<td>het kind dat...</td>
</tr>
<tr>
<td>Conj-Sub</td>
<td>subordinating conjunction</td>
<td>hoezool hij er was</td>
</tr>
</tbody>
</table>
### Figure I–13 Tagging in Dutch (Conj-Sub-Adv to Punct-Slash)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conj-Sub-Adv</td>
<td>interrogative adverb or subordinate</td>
<td>wannen gaat hij weg?</td>
</tr>
<tr>
<td>Det-Art</td>
<td>determiner</td>
<td>een bus</td>
</tr>
<tr>
<td>Det-Dem</td>
<td>demonstrative determiner</td>
<td>deze machine gaat goed</td>
</tr>
<tr>
<td>Det-Indef</td>
<td>indefinite determiner</td>
<td>geen boer</td>
</tr>
<tr>
<td>Det-Int/Rel</td>
<td>interrogative or relative determiner</td>
<td>de vraag uiter man...</td>
</tr>
<tr>
<td>Det-Poss</td>
<td>possessive determiner</td>
<td>mijn boek</td>
</tr>
<tr>
<td>Det-Post/Indef</td>
<td>indefinite postdeterminer</td>
<td>de beide broers</td>
</tr>
<tr>
<td>Det-Pre/Indef</td>
<td>indefinite predeterminer</td>
<td>binnen al deze pakketten</td>
</tr>
<tr>
<td>Inter</td>
<td>interjections</td>
<td>och</td>
</tr>
<tr>
<td>Nm</td>
<td>common noun</td>
<td>boek</td>
</tr>
<tr>
<td>Nm-Net</td>
<td>lowercase and uppercase letters</td>
<td>b, N</td>
</tr>
<tr>
<td>Nm-Net</td>
<td>URL and e-mail address</td>
<td><a href="http://www.inlight.com">www.inlight.com</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:info@inlight.com">info@inlight.com</a></td>
</tr>
<tr>
<td>Num</td>
<td>cardinal number</td>
<td>125, vll, 12/2</td>
</tr>
<tr>
<td>Ord</td>
<td>ordinal number</td>
<td>vijfde, 12ste, 12de</td>
</tr>
<tr>
<td>Part-Inf</td>
<td>particle of Dutch ‘tw infinitive’</td>
<td>hij loopt te gaan</td>
</tr>
<tr>
<td>Part-Neg</td>
<td>negative particle</td>
<td>hij gaat niet snel</td>
</tr>
<tr>
<td>Part-Prefix</td>
<td>separated prefix of (pronominal)</td>
<td>hij loopt mee</td>
</tr>
<tr>
<td>Prep</td>
<td>preposition</td>
<td>in</td>
</tr>
<tr>
<td>Prep-Circ</td>
<td>right part of circumposition</td>
<td>tot nu toe</td>
</tr>
<tr>
<td>Prep-Post</td>
<td>postposition</td>
<td>vele passantes langs komen</td>
</tr>
<tr>
<td>Prep-von</td>
<td>preposition van</td>
<td>van</td>
</tr>
<tr>
<td>Pron-Dem</td>
<td>demonstrative pronoun</td>
<td>deze gaat goed</td>
</tr>
<tr>
<td>Pron-Indef</td>
<td>indefinite pronoun</td>
<td>beide</td>
</tr>
<tr>
<td>Pron-Int/Rel</td>
<td>interrogative or relative pronoun</td>
<td>de vraag wie ...</td>
</tr>
<tr>
<td>Pron-Pers</td>
<td>personal pronoun</td>
<td>hij</td>
</tr>
<tr>
<td>Pron-Rel</td>
<td>relative pronoun</td>
<td>de mon die laichte</td>
</tr>
<tr>
<td>Prop</td>
<td>proper noun, including initials and title of address</td>
<td>Peter, C. Prof.</td>
</tr>
<tr>
<td>Prop-Art</td>
<td>article beginning a name</td>
<td>De Vries</td>
</tr>
<tr>
<td>Prop-Prep</td>
<td>preposition beginning a name</td>
<td>Van den Broek</td>
</tr>
<tr>
<td>Punct</td>
<td>miscellaneous punctuation</td>
<td>{ } [ ] - ---</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>comma</td>
<td>,</td>
</tr>
<tr>
<td>Punct-Quote</td>
<td>quotation type punctuation (includes parentheses)</td>
<td>&quot; &quot; ( )</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>sentence final punctuation</td>
<td>. ! :</td>
</tr>
<tr>
<td>Punct-Slash</td>
<td>slash mark</td>
<td>/</td>
</tr>
</tbody>
</table>

### Figure I–14 Tagging in Dutch (V-Fin to V-PrPart)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>V-Fin</td>
<td>finite verb</td>
<td>zegt</td>
</tr>
<tr>
<td>V-Inf</td>
<td>infinitive verb</td>
<td>zeggen</td>
</tr>
<tr>
<td>V-PaPart</td>
<td>past participle verb</td>
<td>aangepakt</td>
</tr>
<tr>
<td>V-PrPart</td>
<td>present participle verb</td>
<td>zeggend</td>
</tr>
</tbody>
</table>
Tagging in English

The charts in this section provide a reference for tagging in English.

**Figure I–15  Tagging in English (Abbr to Det-Indef)**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbr</td>
<td>abbreviation</td>
<td>i.e.</td>
</tr>
<tr>
<td>Adj</td>
<td>adjective</td>
<td>big</td>
</tr>
<tr>
<td>Adj-Comp</td>
<td>comparative adjective</td>
<td>bigger</td>
</tr>
<tr>
<td>Adj-Ord</td>
<td>ordinal adjective</td>
<td>third</td>
</tr>
<tr>
<td>Adv-Sup</td>
<td>superlative adjective</td>
<td>biggest</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb</td>
<td>quickly</td>
</tr>
<tr>
<td>Adv-Comp</td>
<td>comparative adverb</td>
<td>sooner</td>
</tr>
<tr>
<td>Adv-Int/Ral</td>
<td><em>wh</em>-adverb</td>
<td>how</td>
</tr>
<tr>
<td>Adv-Sup</td>
<td>superlative adverb</td>
<td>soonest</td>
</tr>
<tr>
<td>Aux</td>
<td>auxiliary or modal</td>
<td>could</td>
</tr>
<tr>
<td>Conj-Coord</td>
<td>coordinating conjunction</td>
<td>and</td>
</tr>
<tr>
<td>Conj-Sub</td>
<td>subordinating conjunction</td>
<td>unless</td>
</tr>
<tr>
<td>Det</td>
<td>invariant determiner (singular or plural)</td>
<td>some food</td>
</tr>
<tr>
<td>Det-Def</td>
<td>definite determiner</td>
<td>the</td>
</tr>
<tr>
<td>Det-Indef</td>
<td>indefinite determiner</td>
<td>an</td>
</tr>
</tbody>
</table>
### Figure I–16  Tagging in English (Det-Int to V-Past-have)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Det-Int</td>
<td>interrogative determiner</td>
<td>what time?</td>
</tr>
<tr>
<td>Det-Int/Rel</td>
<td>interrogative or relative determiner</td>
<td>whose</td>
</tr>
<tr>
<td>Det-Pl</td>
<td>plural determiner</td>
<td>those apples</td>
</tr>
<tr>
<td>Det-Poss</td>
<td>possessive determiner</td>
<td>my</td>
</tr>
<tr>
<td>Det-Rel</td>
<td>relative determiner</td>
<td>whatsoever</td>
</tr>
<tr>
<td>Det-Sg</td>
<td>singular determiner</td>
<td>every</td>
</tr>
<tr>
<td>Interj</td>
<td>interjection</td>
<td>oh, hello</td>
</tr>
<tr>
<td>Nn</td>
<td>invariant noun</td>
<td>sheep</td>
</tr>
<tr>
<td>Nn-Letter</td>
<td>letter</td>
<td>b, N</td>
</tr>
<tr>
<td>Nn-Net</td>
<td>URL, e-mail address</td>
<td><a href="http://www.imought.com">www.imought.com</a>,</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:info@imought.com">info@imought.com</a></td>
</tr>
<tr>
<td>Nn-Pl</td>
<td>plural noun</td>
<td>computers</td>
</tr>
<tr>
<td>Nn-Sg</td>
<td>singular noun</td>
<td>farmer</td>
</tr>
<tr>
<td>Num</td>
<td>cardinal number or other numeric expression</td>
<td>40.5, 11/27/00, $12.55, 12%, xxii, 9:00</td>
</tr>
<tr>
<td>Part-Int</td>
<td>infinitive marker</td>
<td>to be or not to be</td>
</tr>
<tr>
<td>Part-Neg</td>
<td>negative particle</td>
<td>not</td>
</tr>
<tr>
<td>Part-Poss</td>
<td>possessive marker</td>
<td>John's coat</td>
</tr>
<tr>
<td>Prep</td>
<td>preposition</td>
<td>below</td>
</tr>
<tr>
<td>Prep-at</td>
<td>preposition at</td>
<td>at</td>
</tr>
<tr>
<td>Prep-of</td>
<td>preposition of</td>
<td>of</td>
</tr>
<tr>
<td>Pron</td>
<td>pronoun</td>
<td>he</td>
</tr>
<tr>
<td>Pron-Inf</td>
<td>wh pronoun</td>
<td>what do you want?</td>
</tr>
<tr>
<td>Pron-Inf/Rel</td>
<td>wh pronoun</td>
<td>who</td>
</tr>
<tr>
<td>Pron-Refl</td>
<td>reflexive pronoun</td>
<td>himself</td>
</tr>
<tr>
<td>Pron-Rel</td>
<td>relative pronoun</td>
<td>whoever</td>
</tr>
<tr>
<td>Prop</td>
<td>name of a person or thing</td>
<td>Grobelard</td>
</tr>
<tr>
<td>Punct</td>
<td>other punctuation</td>
<td>–; /%$</td>
</tr>
<tr>
<td>Punct-Close</td>
<td>closing punctuation</td>
<td>) ) )</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>comma</td>
<td>,</td>
</tr>
<tr>
<td>Punct-Open</td>
<td>opening punctuation</td>
<td>( ( (</td>
</tr>
<tr>
<td>Punct-Quote</td>
<td>quote</td>
<td>“ “</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>sentence-ending punctuation</td>
<td>. . ! ?</td>
</tr>
<tr>
<td>V-Inf-be</td>
<td>infinitive to be</td>
<td>be</td>
</tr>
<tr>
<td>V-Pst-be</td>
<td>verb, past participle, -be verb form</td>
<td>has walked</td>
</tr>
<tr>
<td>V-PstPart-be</td>
<td>past participle of to be</td>
<td>has been</td>
</tr>
<tr>
<td>V-PstPart-have</td>
<td>past participle of to have</td>
<td>has had</td>
</tr>
<tr>
<td>V-Past</td>
<td>verb, past tense</td>
<td>ran</td>
</tr>
<tr>
<td>V-Pst-have</td>
<td>past tense of have</td>
<td>had</td>
</tr>
</tbody>
</table>
**Tagging in Farsi**

The charts in this section provide a reference for tagging in Farsi.

**Figure I–18 Tagging in Farsi (Abbr to Adj-NPB)**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbr</td>
<td>Abbreviation</td>
<td>هن، اوریا</td>
</tr>
<tr>
<td>Acc</td>
<td>Accusative marking on direct objects</td>
<td>نورگر</td>
</tr>
<tr>
<td>Adj</td>
<td>Singular adjective</td>
<td>نورگر</td>
</tr>
<tr>
<td>Adj-Comp</td>
<td>Comparative singular adjective</td>
<td>نورگر</td>
</tr>
<tr>
<td>Adj-Comp-NPB</td>
<td>Singular adjective, marks NP boundary</td>
<td>نورگر</td>
</tr>
<tr>
<td>Adj-Cop</td>
<td>Singular adjective, with copula verb</td>
<td>نورگر</td>
</tr>
<tr>
<td>Adj-Cop-Comp</td>
<td>Singular adjective, with copula verb</td>
<td>نورگر</td>
</tr>
<tr>
<td>Adj-Cop-Sup</td>
<td>Singular adjective, with copula verb</td>
<td>نورگر</td>
</tr>
<tr>
<td>Adj-NPB</td>
<td>Singular adjective, marks NP boundary</td>
<td>نورگر</td>
</tr>
</tbody>
</table>
### Tagging in Farsi (Adj-NoNPB to Adj-Pron-Cop-Sup)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj-NoNPB</td>
<td>Singular adjective, not an NP boundary</td>
<td>زیادی</td>
</tr>
<tr>
<td>Adj-PI</td>
<td>Plural adjective</td>
<td>برگردن</td>
</tr>
<tr>
<td>Adj-PI-Comp</td>
<td>Comparative plural adjective</td>
<td>دیگری</td>
</tr>
<tr>
<td>Adj-PI-Comp-NPB</td>
<td>Comparative plural adjective, marks NP boundary</td>
<td>زیاترا</td>
</tr>
<tr>
<td>Adj-PI-Comp-NoNPB</td>
<td>Comparative plural adjective, not an NP boundary</td>
<td>زیاتراهایی</td>
</tr>
<tr>
<td>Adj-PI-Cop</td>
<td>Plural adjective, with copula verb</td>
<td>ایجادی</td>
</tr>
<tr>
<td>Adj-PI-Cop-Comp</td>
<td>Comparative plural adjective, with copula verb</td>
<td>برگردنی</td>
</tr>
<tr>
<td>Adj-PI-Cop-Sup</td>
<td>Superlative plural adjective, with copula verb</td>
<td>برگردنی‌تر</td>
</tr>
<tr>
<td>Adj-PI-NPB</td>
<td>Plural adjective, marks NP boundary</td>
<td>زیادا</td>
</tr>
<tr>
<td>Adj-PI-NoNPB</td>
<td>Plural adjective, not an NP boundary</td>
<td>زیادهای</td>
</tr>
<tr>
<td>Adj-PI-Pron-Comp-NPB</td>
<td>Comp. plural adjective, with pronoun clitic; NP boundary</td>
<td>کوچک‌ترهایان</td>
</tr>
<tr>
<td>Adj-PI-Pron-Cop</td>
<td>Plural adjective, with clitic and copula verb</td>
<td>پهلوانهایی</td>
</tr>
<tr>
<td>Adj-PI-Pron-Cop-Comp</td>
<td>Comp. plural adjective, with clitic and copula verb</td>
<td>برگردنی‌ترهایی</td>
</tr>
<tr>
<td>Adj-PI-Pron-Cop-Sup</td>
<td>Sup. plural adjective, with clitic and copula verb</td>
<td>ازاسب‌هایی</td>
</tr>
<tr>
<td>Adj-PI-Pron-NPB</td>
<td>Plural adjective, with pronoun clitic; NP boundary</td>
<td>کوچک‌ترهایان</td>
</tr>
<tr>
<td>Adj-PI-Pron-Sup-NPB</td>
<td>Sup. plural adjective, with pronoun clitic; NP boundary</td>
<td>کوچک‌ترهایان</td>
</tr>
<tr>
<td>Adj-PI-Sup</td>
<td>Superlative plural adjective</td>
<td>کوچک‌ترهای</td>
</tr>
<tr>
<td>Adj-PI-Sup-NPB</td>
<td>Superlative plural adjective, marks NP boundary</td>
<td>کوچک‌ترهای</td>
</tr>
<tr>
<td>Adj-PI-Sup-NoNPB</td>
<td>Superlative plural adjective, not an NP boundary</td>
<td>کوچک‌ترهای</td>
</tr>
<tr>
<td>Adj-Pron-Comp-NPB</td>
<td>Comp. singular adjective, with pronoun clitic; NP boundary</td>
<td>کوچک‌ترهای</td>
</tr>
<tr>
<td>Adj-Pron-Cop</td>
<td>Singular adjective, with pronoun clitic and copula verb</td>
<td>کوچک‌ترهای</td>
</tr>
<tr>
<td>Adj-Pron-Cop-Comp</td>
<td>Comp. sing adjective, with pronoun clitic and copula verb</td>
<td>کوچک‌ترهای</td>
</tr>
<tr>
<td>Adj-Pron-Cop-Sup</td>
<td>Sup. sing adjective, with pronoun clitic and copula verb</td>
<td>کوچک‌ترهای</td>
</tr>
<tr>
<td>Adj-Pron-NPB</td>
<td>Singular adjective, with pronoun clitic; NP boundary</td>
<td>زیادهای</td>
</tr>
</tbody>
</table>

---

AUTO LEXER Parts-of-Speech Tagging I-15
<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj-Pron-Sup-NPB</td>
<td>Sup. singular adjective, with pronoun cletic; NP boundary</td>
<td>کوچک‌ترین چندان</td>
</tr>
</tbody>
</table>
Tagging in Finnish

The charts in this section provide a reference for tagging in Finnish.
### Figure I–22 Tagging in Finnish (Abbr to Num-Ord-Loc)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbr</td>
<td>abbreviation</td>
<td>kHz</td>
</tr>
<tr>
<td>Adj</td>
<td>adjective</td>
<td>vaaralleena</td>
</tr>
<tr>
<td>Adj-Gen</td>
<td>genitive adjective</td>
<td>kunnakantarion, saitojen</td>
</tr>
<tr>
<td>Adj-Loc</td>
<td>local adjective</td>
<td>taluisilla, määräk適合isa</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb (including -s derivations)</td>
<td>taso, suuresti, ylihkoemmena</td>
</tr>
<tr>
<td>Adv-Comp-Part</td>
<td>compound part</td>
<td>minimums ja nällintyöskön</td>
</tr>
<tr>
<td>Con-Coord</td>
<td>coordinating conjunction</td>
<td>ja, tai</td>
</tr>
<tr>
<td>Con-Sub</td>
<td>subordinating conjunction</td>
<td>vaikka</td>
</tr>
<tr>
<td>Inte</td>
<td>interjection</td>
<td>hän, ah</td>
</tr>
<tr>
<td>Nn</td>
<td>noun</td>
<td>kylttiön</td>
</tr>
<tr>
<td>Nn-LETTER</td>
<td>lowercase and uppercase letters</td>
<td>b, N</td>
</tr>
<tr>
<td>Nn-Net</td>
<td>URL, e-mail address</td>
<td><a href="http://www.insight.com">www.insight.com</a>, <a href="mailto:info@insight.com">info@insight.com</a></td>
</tr>
<tr>
<td>Nn-Gen</td>
<td>genitive noun</td>
<td>yksikön, ääten</td>
</tr>
<tr>
<td>Nn-LOC</td>
<td>local noun</td>
<td>koivin</td>
</tr>
<tr>
<td>Nn-NOM</td>
<td>nominative noun</td>
<td>maahan</td>
</tr>
<tr>
<td>Nn-Ptv</td>
<td>partitive noun</td>
<td>rikkipiteautta, naarmuja</td>
</tr>
<tr>
<td>Num</td>
<td>numeric expression</td>
<td>1/6, 21.2.1997</td>
</tr>
<tr>
<td>Num-Card</td>
<td>cardinal number</td>
<td>kuudesksi, 6</td>
</tr>
<tr>
<td>Num-Card-Gen</td>
<td>genitive cardinal number</td>
<td>kahdien, 8:n, tuhannien</td>
</tr>
<tr>
<td>Num-Card-Loc</td>
<td>local cardinal number</td>
<td>14:15, viisiosa</td>
</tr>
<tr>
<td>Num-Card-Nom</td>
<td>nominative cardinal number</td>
<td>milloon, 14:15</td>
</tr>
<tr>
<td>Num-Card-Ptv</td>
<td>partitive cardinal number</td>
<td>kolmea, 3:a, 5:tiä, tuhannia</td>
</tr>
<tr>
<td>Num-Gen</td>
<td>other genitive number</td>
<td>kuudennen</td>
</tr>
<tr>
<td>Num-Loc</td>
<td>other local number</td>
<td>kolmesta</td>
</tr>
<tr>
<td>Num-Nom</td>
<td>other nominative number</td>
<td>kahdeksen</td>
</tr>
<tr>
<td>Num-Ord</td>
<td>ordinal number</td>
<td>4. kolmanteena</td>
</tr>
<tr>
<td>Num-Ord-Gen</td>
<td>ordinal genitive number</td>
<td>ensimmäinen, 400:nneen</td>
</tr>
<tr>
<td>Num-Ord-Loc</td>
<td>ordinal local number</td>
<td>4:nneella, biloonareilla</td>
</tr>
</tbody>
</table>
Figure I–23  Tagging in Finnish (Num-Ord-Nom to V-Part)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num-Ord-Nom</td>
<td>ordinal nominative number</td>
<td>kahdeksas, sadan, 100:nhet</td>
</tr>
<tr>
<td>Num-Ord-Plv</td>
<td>ordinal partitive number</td>
<td>kuudesti, 13:tte, sadante, 10:näät</td>
</tr>
<tr>
<td>Num-Plv</td>
<td>partitive number</td>
<td>&quot;kolimatta&quot;</td>
</tr>
<tr>
<td>Prep</td>
<td>either preposition or postposition</td>
<td>itä</td>
</tr>
<tr>
<td>Prep-Post</td>
<td>postposition</td>
<td>osti, kanssa</td>
</tr>
<tr>
<td>Pron</td>
<td>pronoun</td>
<td>minun, sinukai</td>
</tr>
<tr>
<td>Pron-Acc</td>
<td>accusative personal pronoun</td>
<td>minet, hänet, mendät, teidät</td>
</tr>
<tr>
<td>Pron-Gen</td>
<td>genitive pronoun</td>
<td>minun, miniki</td>
</tr>
<tr>
<td>Pron-Loc</td>
<td>local pronoun</td>
<td>tällä monessa, useille</td>
</tr>
<tr>
<td>Pron-Nom</td>
<td>nominative pronoun</td>
<td>minä, eräs, toiset</td>
</tr>
<tr>
<td>Pron-Plv</td>
<td>partitive pronoun</td>
<td>tänä, utes, kumpia</td>
</tr>
<tr>
<td>Prop</td>
<td>proper noun</td>
<td>Markkinaisten</td>
</tr>
<tr>
<td>Prop-Gen</td>
<td>genitive proper noun</td>
<td>Kennedyjen</td>
</tr>
<tr>
<td>Prop-Loc</td>
<td>local proper noun</td>
<td>Helsingissä</td>
</tr>
<tr>
<td>Prop-Nom</td>
<td>nominative proper noun</td>
<td>Kalifornian</td>
</tr>
<tr>
<td>Prop-Plv</td>
<td>partitive proper noun</td>
<td>Vaasana, Virtanen</td>
</tr>
<tr>
<td>Punct</td>
<td>punctuation</td>
<td>; : ; - - - - - - - - -</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>comma</td>
<td>, , , , , , , , , , ,</td>
</tr>
<tr>
<td>Punct-Paren</td>
<td>parenthesis</td>
<td>( ( ( ) ) )</td>
</tr>
<tr>
<td>Punct-Quote</td>
<td>quote</td>
<td>&quot; &quot; &quot; &quot; &quot; &quot; &quot; &quot; &quot; &quot;</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>sentence terminating punctuation</td>
<td>. ! ?</td>
</tr>
<tr>
<td>V-Fin</td>
<td>finite verb</td>
<td>tarvitseen, laitosimeen</td>
</tr>
<tr>
<td>V-Inf</td>
<td>infinitival verb</td>
<td>kirjoittaa, sopiminen</td>
</tr>
<tr>
<td>V-Part</td>
<td>participle</td>
<td>koolitenut, sovittava</td>
</tr>
</tbody>
</table>

**Tagging in French**

The charts in this section provide a reference for tagging in French.
### Figure I–24  Tagging in French (Adj to Num)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj</td>
<td>invariant adjective</td>
<td>heureux</td>
</tr>
<tr>
<td>Adj-Ord-Pl</td>
<td>spelled out plural ordinal number</td>
<td>deuxiémes</td>
</tr>
<tr>
<td>Adj-Ord-Sg</td>
<td>spelled out singular ordinal number</td>
<td>deuxième</td>
</tr>
<tr>
<td>Adj-Pl</td>
<td>plural adjective</td>
<td>gantilles</td>
</tr>
<tr>
<td>Adj-Pre</td>
<td>preposed invariant adjective</td>
<td>vieux</td>
</tr>
<tr>
<td>Adj-Pre-Pl</td>
<td>preposed plural adjective</td>
<td>petits chiens</td>
</tr>
<tr>
<td>Adj-Pre-Sg</td>
<td>preposed singular adjective</td>
<td>petit chien</td>
</tr>
<tr>
<td>Adj-Sg</td>
<td>singular adjective</td>
<td>gantile</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb</td>
<td>finalement aujourd'hui</td>
</tr>
<tr>
<td>Adv-Deg</td>
<td>adverb that can modify an adjective</td>
<td>très</td>
</tr>
<tr>
<td>Aux-Fin-12</td>
<td>1st or 2nd person auxiliary, any tense</td>
<td>fusse</td>
</tr>
<tr>
<td>Aux-Fin-3-Pl</td>
<td>3rd person plural auxiliary, any tense</td>
<td>seraient</td>
</tr>
<tr>
<td>Aux-Fin-3-Sg</td>
<td>3rd person singular auxiliary, any tense</td>
<td>serait</td>
</tr>
<tr>
<td>Aux-Inf</td>
<td>infinitive auxiliary</td>
<td>en avoir assez</td>
</tr>
<tr>
<td>Aux-PsPart</td>
<td>past participle auxiliary</td>
<td>eu</td>
</tr>
<tr>
<td>Aux-PtPart</td>
<td>present participle auxiliary</td>
<td>avant</td>
</tr>
<tr>
<td>Conj-Adv</td>
<td>connecting or subordinating conjunction</td>
<td>quand</td>
</tr>
<tr>
<td>Conj-comme</td>
<td>comme</td>
<td>comme</td>
</tr>
<tr>
<td>Conj-Coord</td>
<td>coordinating conjunction</td>
<td>et, ou</td>
</tr>
<tr>
<td>Conj-que</td>
<td>que</td>
<td>que</td>
</tr>
<tr>
<td>Det-Pl</td>
<td>plural determiner</td>
<td>vos</td>
</tr>
<tr>
<td>Det-Sg</td>
<td>singular determiner</td>
<td>ma, votre</td>
</tr>
<tr>
<td>Interj</td>
<td>interjection</td>
<td>lui, pouce</td>
</tr>
<tr>
<td>Nn</td>
<td>invariant noun</td>
<td>taux</td>
</tr>
<tr>
<td>Nn-Letter</td>
<td>letter</td>
<td>z, K</td>
</tr>
<tr>
<td>Nn-Net</td>
<td>URL and e-mail address</td>
<td><a href="http://www.inxight.com">www.inxight.com</a></td>
</tr>
<tr>
<td>Nn-Pl</td>
<td>plural noun</td>
<td>chiens, fourmis</td>
</tr>
<tr>
<td>Nn-Sg</td>
<td>singular noun</td>
<td>chien, fourmi</td>
</tr>
<tr>
<td>Num</td>
<td>numeral or digit expression</td>
<td>treize, un million, 12.15kHz, XIX</td>
</tr>
</tbody>
</table>
**Figure I–25  Tagging in French (Part-Neg to V-PrPart)**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part-Neg</td>
<td>negation particle</td>
<td>ne</td>
</tr>
<tr>
<td>Part-voi/la</td>
<td>particles voici and voilà</td>
<td>voici</td>
</tr>
<tr>
<td>Prep</td>
<td>proposition (other than à au de du ...)</td>
<td>dans</td>
</tr>
<tr>
<td>Prep-a</td>
<td>proposition à</td>
<td>à, au, aux</td>
</tr>
<tr>
<td>Prep-de</td>
<td>proposition de</td>
<td>de, des, du, d’</td>
</tr>
<tr>
<td>Prep-en</td>
<td>proposition en</td>
<td>en bonne sorte</td>
</tr>
<tr>
<td>Pron</td>
<td>pronoun</td>
<td>il, elles</td>
</tr>
<tr>
<td>Pron-12</td>
<td>1st or 2nd person pronoun</td>
<td>je</td>
</tr>
<tr>
<td>Pron-Clit</td>
<td>clitic pronoun</td>
<td>-lui, -le</td>
</tr>
<tr>
<td>Pron-InterRel</td>
<td>relative or interrogative pronoun (except que)</td>
<td>quoi, quel</td>
</tr>
<tr>
<td>Prop</td>
<td>proper noun</td>
<td>Marie, Paris</td>
</tr>
<tr>
<td>Punct</td>
<td>punctuation</td>
<td>: ;</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>comma</td>
<td>,</td>
</tr>
<tr>
<td>Punct-Quote</td>
<td>quotation marks</td>
<td>“</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>sentence-ending punctuation</td>
<td>. ! ?;</td>
</tr>
<tr>
<td>V/Adj-PaPart</td>
<td>invariant past participial verb or adjective</td>
<td>terminer</td>
</tr>
<tr>
<td>V/Adj-PaPart-Pl</td>
<td>plural past participial verb or adjective</td>
<td>terminez, terminez</td>
</tr>
<tr>
<td>V/Adj-PaPart-Sg</td>
<td>singular past participial verb or adjective</td>
<td>termé, termés, terminé</td>
</tr>
<tr>
<td>V-Fin-12</td>
<td>1st or 2nd person verb, any tense</td>
<td>dites, dites</td>
</tr>
<tr>
<td>V-Fin-3-Pl</td>
<td>3rd person plural verbs, any tense</td>
<td>ditesentent</td>
</tr>
<tr>
<td>V-Fin-3-Sg</td>
<td>3rd person singular verb, any tense</td>
<td>ditesent</td>
</tr>
<tr>
<td>V-Infl</td>
<td>infinitive verb</td>
<td>dines, dines</td>
</tr>
<tr>
<td>V-PrPart</td>
<td>present participle verb</td>
<td>nitant</td>
</tr>
</tbody>
</table>

**Tagging in German**

The charts in this section provide a reference for tagging in German.
<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj-Attr</td>
<td>attributive adjective</td>
<td>schwarze Katze</td>
</tr>
<tr>
<td>Adj-Attr-Comp</td>
<td>comparative attributive adjective</td>
<td>kleinere Datei</td>
</tr>
<tr>
<td>Adj-Attr-Ord</td>
<td>spelled out ordinal numbers</td>
<td>dritter</td>
</tr>
<tr>
<td>Adj-Attr-Sup</td>
<td>superlative attributive adjective</td>
<td>schnellster Läufer</td>
</tr>
<tr>
<td>Adj-Dem</td>
<td>demonstrative adjective</td>
<td>solche</td>
</tr>
<tr>
<td>Adj-Indit</td>
<td>indefinite adjective</td>
<td>deinige</td>
</tr>
<tr>
<td>Adj-Pred</td>
<td>predicate adjective (may also be an adverb)</td>
<td>schnell</td>
</tr>
<tr>
<td>Adj-Pred-Comp</td>
<td>comparative predicate adjective</td>
<td>schneller</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb</td>
<td>immer, zwar</td>
</tr>
<tr>
<td>Adv-int/Rel</td>
<td>interrogative or relative adverb</td>
<td>vieles</td>
</tr>
<tr>
<td>Adv-Pron</td>
<td>pronominal adverb</td>
<td>hierfür</td>
</tr>
<tr>
<td>Aux-fin</td>
<td>finite auxiliary verb</td>
<td>bist</td>
</tr>
<tr>
<td>Aux-inf</td>
<td>auxiliary verb infinitive</td>
<td>gelernt haben</td>
</tr>
<tr>
<td>Aux-PaPart</td>
<td>auxiliary verb past participle</td>
<td>gehabt</td>
</tr>
<tr>
<td>Comp-Left</td>
<td>left compound part</td>
<td>zwei-Kontaktsch.</td>
</tr>
<tr>
<td>Conj/Adv</td>
<td>conjunction or adverb</td>
<td>jedoch</td>
</tr>
<tr>
<td>Conj-als</td>
<td>conjunction als</td>
<td>als</td>
</tr>
<tr>
<td>Conj-Coord</td>
<td>coordinating conjunction</td>
<td>und</td>
</tr>
<tr>
<td>Conj-inf</td>
<td>ininitival conjunction</td>
<td>an ... zu</td>
</tr>
<tr>
<td>Conj-Post</td>
<td>correlative conjunction</td>
<td>wieder ... noch</td>
</tr>
<tr>
<td>Conj-Pre</td>
<td>proposed conjunction</td>
<td>wieder</td>
</tr>
<tr>
<td>Conj-Subord</td>
<td>subordinating conjunction</td>
<td>weil</td>
</tr>
<tr>
<td>Conj-ws</td>
<td>conjunction wie</td>
<td>wie</td>
</tr>
<tr>
<td>DetPron-Quant</td>
<td>quantifying determiner or pronoun</td>
<td>lauter Sachen</td>
</tr>
<tr>
<td>Det-Art</td>
<td>article</td>
<td>die, das</td>
</tr>
<tr>
<td>Det-Dem</td>
<td>demonstrative determiner</td>
<td>diese</td>
</tr>
<tr>
<td>Det-Indit</td>
<td>indefinite determiner</td>
<td>keiner</td>
</tr>
<tr>
<td>Det-IntRel</td>
<td>interrogative determiner or pronoun</td>
<td>viehviel</td>
</tr>
<tr>
<td>Det-Poss</td>
<td>possessive determiner</td>
<td>dein</td>
</tr>
</tbody>
</table>
### Figure I–27 Tagging in German (Interj to Punct)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interj</td>
<td>interjection</td>
<td>ach, oh</td>
</tr>
<tr>
<td>Modal-Fin</td>
<td>finite modal verb</td>
<td>darf</td>
</tr>
<tr>
<td>Modal-Inf</td>
<td>modal verb infinitive</td>
<td>gehören sollen</td>
</tr>
<tr>
<td>Modal-PaPart</td>
<td>modal verb past participle</td>
<td>gegangen gегirt</td>
</tr>
<tr>
<td>N-1</td>
<td>noun</td>
<td>Tisch</td>
</tr>
<tr>
<td>N-Letter</td>
<td>lowercase letters with or without a period and uppercase letters</td>
<td>e, p, a.</td>
</tr>
<tr>
<td>N-Net</td>
<td>URL and e-mail address</td>
<td><a href="http://www.inlight.com">www.inlight.com</a>, <a href="mailto:info@inlight.com">info@inlight.com</a></td>
</tr>
<tr>
<td>Num</td>
<td>cardinal number or date</td>
<td>zwei, 2.3.1999</td>
</tr>
<tr>
<td>Num-Ord</td>
<td>ordinal number</td>
<td>43.</td>
</tr>
<tr>
<td>Part-Ant</td>
<td>sentential particle</td>
<td>danke</td>
</tr>
<tr>
<td>Part-Comp</td>
<td>comparative particle</td>
<td>viel besser</td>
</tr>
<tr>
<td>Part-Inf</td>
<td>infinitive particle</td>
<td>zu sagen</td>
</tr>
<tr>
<td>Part-Neg</td>
<td>negation particle</td>
<td>nicht</td>
</tr>
<tr>
<td>Part-Num</td>
<td>numerical particle</td>
<td>rund 50 Dattien</td>
</tr>
<tr>
<td>Part-Pos</td>
<td>positive particle</td>
<td>zu schnell</td>
</tr>
<tr>
<td>Part-Prep</td>
<td>separable prefix</td>
<td>er sprach mich an</td>
</tr>
<tr>
<td>Part-Sup</td>
<td>superlative particle</td>
<td>am besten</td>
</tr>
<tr>
<td>Prep-aus</td>
<td>preposition aus</td>
<td>aus</td>
</tr>
<tr>
<td>Prep-Circ</td>
<td>last part of a conjunction</td>
<td>um Himmels willen</td>
</tr>
<tr>
<td>Prep-Det</td>
<td>preposition-article combination</td>
<td>zum</td>
</tr>
<tr>
<td>Prep-fur</td>
<td>preposition für</td>
<td>für</td>
</tr>
<tr>
<td>Prep-Gegen</td>
<td>postposition</td>
<td>dem Haus gegenüber</td>
</tr>
<tr>
<td>Prep-Pro</td>
<td>preposition pro</td>
<td>mit</td>
</tr>
<tr>
<td>Prep-Pr</td>
<td>preposition pro</td>
<td>pro</td>
</tr>
<tr>
<td>Prep-von</td>
<td>preposition von</td>
<td>von</td>
</tr>
<tr>
<td>Pron-Demo</td>
<td>demonstrative pronoun</td>
<td>diese ist besser</td>
</tr>
<tr>
<td>Pron-Demo-Inv</td>
<td>uninflated demonstrative</td>
<td>soch ein Erfolg</td>
</tr>
<tr>
<td>Pron-Indet</td>
<td>indefinite pronoun</td>
<td>niemand</td>
</tr>
<tr>
<td>Pron-Indet-Inv</td>
<td>uninflated determiner</td>
<td>manch ein Mensch</td>
</tr>
<tr>
<td>Pron-Infl</td>
<td>interrogative or relative pronoun</td>
<td>was, wer</td>
</tr>
<tr>
<td>Pron-Infl-Rel-Inv</td>
<td>uninflated interrogative or relative pronoun</td>
<td>was für</td>
</tr>
<tr>
<td>Pron-Pers</td>
<td>personal pronoun</td>
<td>ich</td>
</tr>
<tr>
<td>Pron-Poss</td>
<td>possessive pronoun</td>
<td>meine sind gelb</td>
</tr>
<tr>
<td>Pron-Rezip</td>
<td>reciprocal pronoun</td>
<td>einander</td>
</tr>
<tr>
<td>Pron-Pers</td>
<td>reflexive pronoun</td>
<td>sich</td>
</tr>
<tr>
<td>Pron-Pers</td>
<td>relative pronoun</td>
<td>die Leute, die...</td>
</tr>
<tr>
<td>Punct</td>
<td>punctuation</td>
<td>(</td>
</tr>
</tbody>
</table>

### Figure I–28 Tagging in German (Punct-Comma to V-PaPart)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punct-Comma</td>
<td>comma</td>
<td>,</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>sentence-ending punctuation</td>
<td>. ! ?</td>
</tr>
<tr>
<td>V-Fin</td>
<td>finite verb</td>
<td>schummert</td>
</tr>
<tr>
<td>V-Inf</td>
<td>infinitive verb</td>
<td>er kann schummern</td>
</tr>
<tr>
<td>V-izu</td>
<td>zu infinitive</td>
<td>auszubilden</td>
</tr>
<tr>
<td>V-PaPart</td>
<td>past participle verb</td>
<td>er ist geschummern</td>
</tr>
</tbody>
</table>
### Tagging in Italian

#### Figure I–29  Tagging in Italian (Adj-ex to Conj-Adv)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj-ex</td>
<td>the prefix ex</td>
<td>ex</td>
</tr>
<tr>
<td>Adj-Pl</td>
<td>plural adjective (includes ordinals: seconds)</td>
<td>belle</td>
</tr>
<tr>
<td>Adj-PtPart-Pl</td>
<td>plural present participle verb</td>
<td>meditanti, destreggianti</td>
</tr>
<tr>
<td>Adj-PtPart-Pl-Pron</td>
<td>plural present participle verb with attached clitics</td>
<td>fascianterella, quietanteri</td>
</tr>
<tr>
<td>Adj-PtPart-Sg</td>
<td>singular present participle verb</td>
<td>meditante, destreggiante</td>
</tr>
<tr>
<td>Adj-PtPart-Sg-Pron</td>
<td>singular present participle verb with attached clitics</td>
<td>spumantello, constanti</td>
</tr>
<tr>
<td>Adj-Sg</td>
<td>singular adjective (includes ordinals: seconds, 2)</td>
<td>buono, narcisstico</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb</td>
<td>funettisticamente</td>
</tr>
<tr>
<td>Aux</td>
<td>finite auxiliary (be and have)</td>
<td>saranno, avrete</td>
</tr>
<tr>
<td>Aux-Ger</td>
<td>gerundive auxiliary</td>
<td>essendo, avendo</td>
</tr>
<tr>
<td>Aux-Impv</td>
<td>imperative auxiliary</td>
<td>sì, abbi</td>
</tr>
<tr>
<td>Aux-Inf</td>
<td>infinitive auxiliary</td>
<td>esser, aver</td>
</tr>
<tr>
<td>Aux-PtPart-Pl</td>
<td>plural past participle auxiliary</td>
<td>avuti, avute</td>
</tr>
<tr>
<td>Aux-PtPart-Sg</td>
<td>singular past participle auxiliary</td>
<td>avuta, avuto</td>
</tr>
<tr>
<td>Auxiliary Pt-Pl</td>
<td>plural present participle auxiliary</td>
<td>essenti, aventi</td>
</tr>
<tr>
<td>Auxiliary PtPart-Sg</td>
<td>singular present participle auxiliary</td>
<td>essente, avente</td>
</tr>
<tr>
<td>Conj</td>
<td>conjunction</td>
<td>tuttavia</td>
</tr>
<tr>
<td>Conjunction</td>
<td>interrogative adverb</td>
<td>quando, dove, come</td>
</tr>
</tbody>
</table>
### Figure I–30  Tagging in Italian (Conj-che to Prep-Pre)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conj-che</td>
<td>the connector <em>che</em></td>
<td>chi, che</td>
</tr>
<tr>
<td>Conj-Coord</td>
<td>coordinating conjunction</td>
<td>ed, e/o</td>
</tr>
<tr>
<td>Conj-Pre</td>
<td>first part of a multword conjunction</td>
<td>dato, che</td>
</tr>
<tr>
<td>Det/Pron-Dem-Pl</td>
<td>plural demonstrative determiner or pronoun</td>
<td>quale</td>
</tr>
<tr>
<td>Det/Pron-Dem-Sg</td>
<td>singular demonstrative determiner or pronoun</td>
<td>costei</td>
</tr>
<tr>
<td>Det/Pron-Int-Pl</td>
<td>plural interrogative determiner or pronoun</td>
<td>quanti soldi</td>
</tr>
<tr>
<td>Det/Pron-Int-Sg</td>
<td>singular interrogative determiner or pronoun</td>
<td>quel, cos'</td>
</tr>
<tr>
<td>Det/Pron-Poss-Pl</td>
<td>plural possessive determiner or pronoun</td>
<td>mie, vostri</td>
</tr>
<tr>
<td>Det/Pron-Poss-Sg</td>
<td>singular possessive determiner or pronoun</td>
<td>nostro, sua</td>
</tr>
<tr>
<td>Det/Pron-Quant</td>
<td>invariant quantifying determiner or pronoun</td>
<td>qualunque, qualcuno</td>
</tr>
<tr>
<td>Det/Pron-Quant-Pl</td>
<td>plural quantifying determiner or pronoun</td>
<td>molti uomini</td>
</tr>
<tr>
<td>Det/Pron-Quant-Sg</td>
<td>singular quantifying determiner or pronoun</td>
<td>molta gente</td>
</tr>
<tr>
<td>Det-PI</td>
<td>plural determiner</td>
<td>qual</td>
</tr>
<tr>
<td>Det-Pre</td>
<td>pre-determiner</td>
<td>tutto il giorno</td>
</tr>
<tr>
<td>Det-Sg</td>
<td>singular determiner</td>
<td>quel</td>
</tr>
<tr>
<td>Interj</td>
<td>interjection or onomatopoeia</td>
<td>uff, perdiana, eh</td>
</tr>
<tr>
<td>Nn-Letter</td>
<td>lowercase and uppercase letters, by themselves or followed by a period or right parenthesis</td>
<td>b, N</td>
</tr>
<tr>
<td>Nn-Net</td>
<td>URL, e-mail address</td>
<td><a href="http://www.inight.com">www.inight.com</a>, <a href="mailto:info@inight.com">info@inight.com</a></td>
</tr>
<tr>
<td>Nn-Pl</td>
<td>plural noun</td>
<td>case</td>
</tr>
<tr>
<td>Nn-Sg</td>
<td>singular noun</td>
<td>casa, bolzano</td>
</tr>
<tr>
<td>Num</td>
<td>numeric expression (in digits)</td>
<td>+5, 23.05, 4,84, 1997</td>
</tr>
<tr>
<td>Prep</td>
<td>preposition</td>
<td>tra, con</td>
</tr>
<tr>
<td>Prep-a</td>
<td>preposition a</td>
<td>a</td>
</tr>
<tr>
<td>Prep-da</td>
<td>preposition da</td>
<td>da</td>
</tr>
<tr>
<td>Prep-Det-Pl</td>
<td>combination preposition and plural determiner</td>
<td>sulle, sugli, pegli</td>
</tr>
<tr>
<td>Prep-Det-Pla</td>
<td>combination a and plural determiner</td>
<td>ai, alle</td>
</tr>
<tr>
<td>Prep-Det-Plda</td>
<td>combination da and plural determiner</td>
<td>dalle</td>
</tr>
<tr>
<td>Prep-Det-Pldi</td>
<td>combination di and pl. determiner</td>
<td>della</td>
</tr>
<tr>
<td>Prep-Det-Sg</td>
<td>combination preposition and singular determiner</td>
<td>sullo, nello</td>
</tr>
<tr>
<td>Prep-Det-Sgra</td>
<td>combination a and sg. determiner</td>
<td>al, allo</td>
</tr>
<tr>
<td>Prep-Det-Sgda</td>
<td>combination da and sg. determiner</td>
<td>dalla</td>
</tr>
<tr>
<td>Prep-Det-Sgdi</td>
<td>combination di and sg. determiner</td>
<td>delli</td>
</tr>
<tr>
<td>Prep-di</td>
<td>preposition <em>di</em></td>
<td>di</td>
</tr>
<tr>
<td>Prep-Pre</td>
<td>first word of a multword preposition</td>
<td>per mezzo</td>
</tr>
</tbody>
</table>
### Figure I–31 Tagging in Italian (Pron to V-Inf-Pron)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pron</td>
<td>invariant pronoun</td>
<td>sé</td>
</tr>
<tr>
<td>Pron-cho</td>
<td>interrogative pronoun chi</td>
<td>chi</td>
</tr>
<tr>
<td>Pron-Clinic</td>
<td>clitic pronoun</td>
<td>vi, no, on, glielo</td>
</tr>
<tr>
<td>Pron-Clinic-Pre</td>
<td>first of a two-clitic sequence</td>
<td>ce, ve</td>
</tr>
<tr>
<td>Pron-Indef-Pl</td>
<td>plural indefinite pronoun</td>
<td>tutti amano la vacca.</td>
</tr>
<tr>
<td>Pron-Indef-Sg</td>
<td>singular indefinite pronoun</td>
<td>qualcuno</td>
</tr>
<tr>
<td>Pron-Pl</td>
<td>plural pronoun</td>
<td>noi</td>
</tr>
<tr>
<td>Pron-Rel</td>
<td>invariant relative pronoun</td>
<td>cui</td>
</tr>
<tr>
<td>Pron-Rel-Pl</td>
<td>plural relative pronoun</td>
<td>i bambini i quali</td>
</tr>
<tr>
<td>Pron-Rel-Sg</td>
<td>singular relative pronoun</td>
<td>il bambino il quale</td>
</tr>
<tr>
<td>Pron-Sg</td>
<td>singular pronoun</td>
<td>lei, lui</td>
</tr>
<tr>
<td>Prop</td>
<td>proper noun</td>
<td>Bernardo, Monte Isola</td>
</tr>
<tr>
<td>Punct</td>
<td>punctuation</td>
<td>; : - ! ? ;</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>comma</td>
<td>,</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>sentence punctuation</td>
<td>. ! ? ;</td>
</tr>
<tr>
<td>V/Adj-PaPart-Pl</td>
<td>plural past participle verb or adjective</td>
<td>riposti, offuscate</td>
</tr>
<tr>
<td>V/Adj-PaPart-Pl-Pron</td>
<td>plural past participle verb or adjective, with attached clinic</td>
<td>telatesele, assestatoci</td>
</tr>
<tr>
<td>V/Adj-PaPart-Sg</td>
<td>singular past participle verb or adjective</td>
<td>sbudellato</td>
</tr>
<tr>
<td>V/Adj-PaPart-Sg-Pron</td>
<td>singular past participle verb or adjective, with attached clinic</td>
<td>commossosi, ingrantiommi</td>
</tr>
<tr>
<td>V-Inf-Pron</td>
<td>finite verb with attached clinic</td>
<td>blisteraremo</td>
</tr>
<tr>
<td>V-Inf-Pron</td>
<td>finite verb with attached clinic</td>
<td>trattare, leggervarsi</td>
</tr>
<tr>
<td>V-Ger</td>
<td>gerund</td>
<td>adducendo, intervetstando</td>
</tr>
<tr>
<td>V-Ger-Pron</td>
<td>gerund with attached clinic</td>
<td>saziandotele, appurandolo</td>
</tr>
<tr>
<td>V-Impr</td>
<td>imperative verb</td>
<td>Va’ a casa!</td>
</tr>
<tr>
<td>V-Impr-Pron</td>
<td>imperative verb with attached clinic</td>
<td>russetti, planaci</td>
</tr>
<tr>
<td>V-Inf</td>
<td>infinitive verb</td>
<td>sciupare, trascinar</td>
</tr>
<tr>
<td>V-Inf-Pron</td>
<td>infinitive verb with attached clinic</td>
<td>spolciare, risucchiare</td>
</tr>
</tbody>
</table>
Tagging in Japanese

**Figure I–32** Tagging in Japanese

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suf</td>
<td>suffix</td>
<td>さん</td>
</tr>
<tr>
<td>Verb</td>
<td>verb</td>
<td>読む、理解する</td>
</tr>
<tr>
<td>Verb-D</td>
<td>verb denoting modality and aspect</td>
<td>はじめめる、できる、（て）いる</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb</td>
<td>ゆっくり、じっと</td>
</tr>
<tr>
<td>Aux</td>
<td>auxiliary verb</td>
<td>だ、です、ない</td>
</tr>
<tr>
<td>Case</td>
<td>case marker</td>
<td>が、を、さえ</td>
</tr>
<tr>
<td>Conj</td>
<td>conjunction</td>
<td>そして、しかし</td>
</tr>
<tr>
<td>Interj</td>
<td>interjection</td>
<td>さあ、えっ</td>
</tr>
<tr>
<td>Nn</td>
<td>noun</td>
<td>先生、分析、ファイル</td>
</tr>
<tr>
<td>Nn-Adv</td>
<td>noun commonly used adverbally</td>
<td>今日、午後、1月</td>
</tr>
<tr>
<td>Nn-Assi</td>
<td>ASCII character, sequence or word</td>
<td>computer</td>
</tr>
<tr>
<td>Nn-D</td>
<td>formal noun, non-contentual noun</td>
<td>こと、の、も、よう</td>
</tr>
<tr>
<td>Nn-Pron</td>
<td>pronoun</td>
<td>あなた、私、ここ</td>
</tr>
<tr>
<td>Num</td>
<td>numeric nominal</td>
<td>山田、富士山</td>
</tr>
<tr>
<td>Pre</td>
<td>nominal prefix</td>
<td>お（水）、高（倍算）</td>
</tr>
<tr>
<td>Punct</td>
<td>punctuation</td>
<td>&quot; : # @</td>
</tr>
<tr>
<td>Punct-Close</td>
<td>closing punctuation</td>
<td>、）、（</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>comma</td>
<td>,</td>
</tr>
<tr>
<td>Punct-Open</td>
<td>opening punctuation</td>
<td>（,《、「</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>sentence-ending punctuation</td>
<td>。、?</td>
</tr>
</tbody>
</table>
**Tagging in Korean**

*Figure I–33 Tagging in Korean (Adv to Pron-Case-Pl-Acc)*

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adv</td>
<td>Adverb</td>
<td>만들, 그리면</td>
</tr>
<tr>
<td>Case</td>
<td>Case Marker</td>
<td>이, 하고</td>
</tr>
<tr>
<td>Det</td>
<td>Determiner</td>
<td>이, 그</td>
</tr>
<tr>
<td>Interj</td>
<td>Interjection</td>
<td>데하고, 해</td>
</tr>
<tr>
<td>Nn</td>
<td>Noun</td>
<td>사자, 포게리</td>
</tr>
<tr>
<td>Nh-Ascii</td>
<td>ASCII characters, Noun</td>
<td>copyright, Web</td>
</tr>
<tr>
<td>Nh-Case</td>
<td>Noun Case Marked</td>
<td>초래율, 두부들</td>
</tr>
<tr>
<td>Nh-Case-Acc</td>
<td>Noun Case Marked Accusative</td>
<td>조각을</td>
</tr>
<tr>
<td>Nh-Case-Conj</td>
<td>Conjoined Case Marked Noun</td>
<td>교열아는, 도래아는</td>
</tr>
<tr>
<td>Nh-Case-Conj-Pl</td>
<td>Conjoined Case Marked Plural Noun</td>
<td>교열숙해주는</td>
</tr>
<tr>
<td>Nh-Case-Dij</td>
<td>Disjunctive Case Marked Noun</td>
<td>여기까지나</td>
</tr>
<tr>
<td>Nh-Case-Dij-Pl</td>
<td>Disjunctive Case Marked Plural Noun</td>
<td>벤자들해가나</td>
</tr>
<tr>
<td>Nh-Case-Gen</td>
<td>Noun Case Marked Genitive</td>
<td>쌤변다</td>
</tr>
<tr>
<td>Nh-Case-PI</td>
<td>Noun Case Marked-Plural</td>
<td>싸우남슬슬, 금성들슬</td>
</tr>
<tr>
<td>Nh-Case-Pl-Acc</td>
<td>Noun Case Marked-Plural Accusative</td>
<td>찢짐될때</td>
</tr>
<tr>
<td>Nh-Case-Pl-Gen</td>
<td>Noun Case Marked-Plural Genitive</td>
<td>교수들임</td>
</tr>
<tr>
<td>Nh-Conj</td>
<td>Conjoined Noun</td>
<td>강이자와, 자기와</td>
</tr>
<tr>
<td>Nh-Conj-PI</td>
<td>Conjoined Plural Noun</td>
<td>교연들하고, 전구들과하고</td>
</tr>
<tr>
<td>Nh-Dij</td>
<td>Disjunctive Noun</td>
<td>겸손이나, 과렴까나</td>
</tr>
<tr>
<td>Nh-Dij-Pl</td>
<td>Disjunctive Plural Noun</td>
<td>힘차들이나, 권장들이나</td>
</tr>
<tr>
<td>Nh-Pl</td>
<td>Noun-Plural</td>
<td>사람들, 해사들</td>
</tr>
<tr>
<td>Nnun</td>
<td>Numeral Expression</td>
<td>30 번, 삼십번</td>
</tr>
<tr>
<td>Pron</td>
<td>Pronoun</td>
<td>나, 너</td>
</tr>
<tr>
<td>Pron-Case</td>
<td>Case Marked Pronoun</td>
<td>나는, 너를</td>
</tr>
<tr>
<td>Pron-Case-Acc</td>
<td>Pronoun Case Marked Accusative</td>
<td>너를</td>
</tr>
<tr>
<td>Pron-Case-Conj</td>
<td>Pronoun Case Marked-Conjunctive</td>
<td>제네해가는</td>
</tr>
<tr>
<td>Pron-Case-Conj-Pl</td>
<td>Pronoun Case Marked-Conjunctive Plural</td>
<td>우리들해가는</td>
</tr>
<tr>
<td>Pron-Case-Dij</td>
<td>Pronoun Case Marked-Disjunctive</td>
<td>여기까지나, 맞을</td>
</tr>
<tr>
<td>Pron-Case-Dij-Pl</td>
<td>Pronoun Case Marked-Disjunctive Plural</td>
<td>저여들까지나</td>
</tr>
<tr>
<td>Pron-Case-Gen</td>
<td>Pronoun Case Marked Genitive</td>
<td>나, 이</td>
</tr>
<tr>
<td>Pron-Case-Pl</td>
<td>Pronoun Case Marked Plural</td>
<td>우리들임</td>
</tr>
<tr>
<td>Pron-Case-Pl-Acc</td>
<td>Pronoun Case Marked Plural Accusative</td>
<td>우리들임</td>
</tr>
</tbody>
</table>
**Figure I–34 Tagging in Korean (Pron-Case-Pl-Gen to V-Pres)**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pron-Case-Pl-Gen</td>
<td>Pronoun Case Marked Plural Genitive</td>
<td>우리들의의</td>
</tr>
<tr>
<td>Pron-Conj</td>
<td>Conjoined Pronoun</td>
<td>자네하고</td>
</tr>
<tr>
<td>Pron-Conj-Pl</td>
<td>Conjoined Plural Pronoun</td>
<td>우리들하고</td>
</tr>
<tr>
<td>Pron-Disj</td>
<td>Disjunctive Pronoun</td>
<td>자기나</td>
</tr>
<tr>
<td>Pron-Disj-Pl</td>
<td>Disjunctive Plural Pronoun</td>
<td>우리들이나</td>
</tr>
<tr>
<td>Pron-Pl</td>
<td>Plural Pronoun</td>
<td>우리들, 그들</td>
</tr>
<tr>
<td>Punct</td>
<td>Punctuation</td>
<td>:;</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>Punctuation-Comma</td>
<td>.</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>Punctuation-Sentence</td>
<td>.</td>
</tr>
<tr>
<td>V-Fut</td>
<td>Future Tense Verb</td>
<td>완매하겠다, 시촉하겠다</td>
</tr>
<tr>
<td>V-Post</td>
<td>Past Tense Verb</td>
<td>졸분했었는, 졸었었다</td>
</tr>
<tr>
<td>V-PreMod</td>
<td>Pre-modifying Verb</td>
<td>졸은</td>
</tr>
<tr>
<td>V-Pres</td>
<td>Present Tense Verb</td>
<td>졸고하려, 범거름다</td>
</tr>
</tbody>
</table>
### Tagging in Bokmal

**Figure I–35 Tagging in Bokmal (Abbr to AuxN-PrPart)**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbr</td>
<td>abbreviation</td>
<td>adv., ibid.</td>
</tr>
<tr>
<td>Adj-Comp</td>
<td>comparative adjective</td>
<td>mindre, eldre</td>
</tr>
<tr>
<td>Adj-Comp-Gen</td>
<td>genitive comparative adjective</td>
<td>elders</td>
</tr>
<tr>
<td>Adj-Def-PI</td>
<td>definite and/or plural adjective</td>
<td>gråne, mørke, kaldé</td>
</tr>
<tr>
<td>Adj-Def-PI-Gen</td>
<td>genitive definite and/or plural adjective</td>
<td>fremmede, uøremes</td>
</tr>
<tr>
<td>Adj-Def-Sg</td>
<td>definite singular adjective</td>
<td>lille</td>
</tr>
<tr>
<td>Adj-Def-Sg-Gen</td>
<td>genitive definite singular adjective</td>
<td>lilles</td>
</tr>
<tr>
<td>Adj-Indef-Sg</td>
<td>indefinite singular adjective</td>
<td>skriftlig</td>
</tr>
<tr>
<td>Adj-Indef-Sg-Gen</td>
<td>genitive indefinite singular adjective</td>
<td>skriftlig</td>
</tr>
<tr>
<td>Adj-PaPart-Def-PI</td>
<td>definite and/or plural adjectival past participle</td>
<td>besøkte, besødte, dempede</td>
</tr>
<tr>
<td>Adj-PaPart-Def-PI-Gen</td>
<td>genitive, definite and/or plural, adjectival past participle</td>
<td>besøkt, besødt, dempedes</td>
</tr>
<tr>
<td>Adj-PaPart-Indef-Sg</td>
<td>adjectival past participle, indefinite singular</td>
<td>for kryptisk kommunikasjon</td>
</tr>
<tr>
<td>Adj-PI</td>
<td>plural adjective</td>
<td>små problemer</td>
</tr>
<tr>
<td>Adj-PI-Gen</td>
<td>genitive plural adjective</td>
<td>smås</td>
</tr>
<tr>
<td>Adj-PrPart</td>
<td>adjectival present participle</td>
<td>begynnelsen, styrrende</td>
</tr>
<tr>
<td>Adj-PrPart-Gen</td>
<td>genitive adjectival present participle</td>
<td>reindenes, laveknes, gråterenes</td>
</tr>
<tr>
<td>Adj-Sup</td>
<td>superlative adjective</td>
<td>best, mest, størst</td>
</tr>
<tr>
<td>Adj-Sup-Def</td>
<td>definite superlative adjective</td>
<td>fineste, innersete, viktigaste</td>
</tr>
<tr>
<td>Adj-Sup-Def-Gen</td>
<td>genitive definite superlative adjective</td>
<td>fineste, innersete, viktigaste</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb</td>
<td>ikke, oppo, heller</td>
</tr>
<tr>
<td>Adv-Comp</td>
<td>comparative adverb</td>
<td>længer</td>
</tr>
<tr>
<td>Adv-Sup</td>
<td>superlative adverb</td>
<td>heller</td>
</tr>
<tr>
<td>AuxV-Impv</td>
<td>imperative auxiliary or main verb</td>
<td>bli</td>
</tr>
<tr>
<td>AuxV-Inf</td>
<td>infinitive auxiliary or main verb</td>
<td>våre</td>
</tr>
<tr>
<td>AuxV-Inf-SForm</td>
<td>S-Form of infinitive auxiliary or main verb</td>
<td>has</td>
</tr>
<tr>
<td>AuxV-PaPart</td>
<td>past participle of auxiliary or main verb</td>
<td>hatt, var, blitt, flitt</td>
</tr>
<tr>
<td>AuxV-Past</td>
<td>past tense auxiliary or main verb</td>
<td>haddé, var, bile</td>
</tr>
<tr>
<td>AuxV-Pr</td>
<td>present tense auxiliary or main verb</td>
<td>har, er, blir, får</td>
</tr>
<tr>
<td>AuxV-PrPart</td>
<td>present participle of auxiliary or main verb</td>
<td>havende, blivende</td>
</tr>
</tbody>
</table>
**Figure I–36  Tagging in Bokmal (Aux-Inf to Num-Indef-Sg)**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aux-Inf</td>
<td>infinitive auxiliary verb</td>
<td>kune, måtte</td>
</tr>
<tr>
<td>Aux-Inf-SForm</td>
<td>S-Form of infinitive auxiliary verb</td>
<td>kunenes, måttene</td>
</tr>
<tr>
<td>Aux-PaPart</td>
<td>past participle of auxiliary verb</td>
<td>kunnet, måttet</td>
</tr>
<tr>
<td>Aux-Past</td>
<td>past tense auxiliary verb</td>
<td>kune, måtte</td>
</tr>
<tr>
<td>Aux-Pres</td>
<td>present tense auxiliary verb</td>
<td>kan, vil</td>
</tr>
<tr>
<td>Aux-Pres-SForm</td>
<td>S-Form of present tense auxiliary verb</td>
<td>villenes, skilles</td>
</tr>
<tr>
<td>Aux-PfPart</td>
<td>present participle of auxiliary verb</td>
<td>villande, skaldande</td>
</tr>
<tr>
<td>Consp-Part</td>
<td>left compound part</td>
<td>kaster- og forretningsbygg</td>
</tr>
<tr>
<td>Conq</td>
<td>subordinating or relativating conjunction</td>
<td>som, mens</td>
</tr>
<tr>
<td>Conq-Coord</td>
<td>coordinating conjunction</td>
<td>og, eller</td>
</tr>
<tr>
<td>Det-Art-Def-Pl</td>
<td>definite plural determiner</td>
<td>disse</td>
</tr>
<tr>
<td>Det-Art-Def-Sg</td>
<td>definite singular determiner</td>
<td>på denne grunn</td>
</tr>
<tr>
<td>Det-Art-Indef</td>
<td>indefinite singular determiner</td>
<td>en, et</td>
</tr>
<tr>
<td>Det-Pron-Comp</td>
<td>determiner or pronoun, comparative</td>
<td>mer</td>
</tr>
<tr>
<td>Det-Pron-Pl</td>
<td>determiner or pronoun, plural</td>
<td>noen, hvite, albo, andre</td>
</tr>
<tr>
<td>Det-Pron-Pl-Gen</td>
<td>determiner or pronoun, plural genitive</td>
<td>noen, alve, ankes</td>
</tr>
<tr>
<td>Det-Pron-Sg</td>
<td>determiner or pronoun, singular</td>
<td>hver, litt, alt</td>
</tr>
<tr>
<td>Det-Pron-Sg-Gen</td>
<td>determiner or pronoun, singular genitive</td>
<td>enlever, anners</td>
</tr>
<tr>
<td>Det-Pron-Sup-Def</td>
<td>determiner or pronoun, definite superlative</td>
<td>mest</td>
</tr>
<tr>
<td>Inter</td>
<td>interjection</td>
<td>ja, herregd</td>
</tr>
<tr>
<td>Nin-Def-Pl</td>
<td>definite plural noun</td>
<td>dorren, armene</td>
</tr>
<tr>
<td>Nin-Def-Pl-Gen</td>
<td>definite definite plural noun</td>
<td>salongenes, kollegnes</td>
</tr>
<tr>
<td>Nin-Def-Sg</td>
<td>definite singular noun</td>
<td>flyet, stokken</td>
</tr>
<tr>
<td>Nin-Def-Sg-Gen</td>
<td>definite definite singular noun</td>
<td>selskapets, spisesauns</td>
</tr>
<tr>
<td>Nin-Indef-Pl</td>
<td>indefinite plural noun</td>
<td>plaser,Mapper</td>
</tr>
<tr>
<td>Nin-Indef-Pl-Gen</td>
<td>indefinite definite plural noun</td>
<td>tiders, menneekser</td>
</tr>
<tr>
<td>Nin-Indef-Sg</td>
<td>indefinite singular noun</td>
<td>stof, stripe</td>
</tr>
<tr>
<td>Nin-Indef-Sg-Gen</td>
<td>indefinite definite singular noun</td>
<td>topps</td>
</tr>
<tr>
<td>Nin-Indef-SP</td>
<td>indefinite singular or plural noun</td>
<td>lys, skritt</td>
</tr>
<tr>
<td>Nin-Indef-SP-Gen</td>
<td>indefinite singular or plural noun</td>
<td>slags, lands, års</td>
</tr>
<tr>
<td>Nin-Letter</td>
<td>lowercase and uppercase letters</td>
<td>b, N</td>
</tr>
<tr>
<td>Nin-Net</td>
<td>URL and e-mail address</td>
<td><a href="http://www.might.com">www.might.com</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:info@might.com">info@might.com</a></td>
</tr>
<tr>
<td>Num</td>
<td>cardinal numeric expression or plural cardinal number (spelled out)</td>
<td>-294, 4.6%, xii, 1.100</td>
</tr>
<tr>
<td>Num-Def-Sg</td>
<td>the number &quot;one&quot;, def. sg. (spelled out)</td>
<td>en</td>
</tr>
<tr>
<td>Num-Indef-Sg</td>
<td>the number &quot;one&quot;, indef. sg. (spelled out)</td>
<td>en, ett</td>
</tr>
</tbody>
</table>
### Figure I–37  Tagging in Bokmal (Ord to V-PrPart)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ord</td>
<td>ordinal number (in digits or spelled out)</td>
<td>7., første</td>
</tr>
<tr>
<td>Part-Inf</td>
<td>infinitival particle</td>
<td>å beskrive</td>
</tr>
<tr>
<td>Prep</td>
<td>proposition</td>
<td>med, ut</td>
</tr>
<tr>
<td>Prep-av</td>
<td>proposition av</td>
<td>av</td>
</tr>
<tr>
<td>Prep-for</td>
<td>proposition for</td>
<td>for</td>
</tr>
<tr>
<td>Prep-fra</td>
<td>proposition fra</td>
<td>fra</td>
</tr>
<tr>
<td>Prep-i</td>
<td>proposition i</td>
<td>i</td>
</tr>
<tr>
<td>Prep-opp</td>
<td>proposition opp</td>
<td>på bakgrunn</td>
</tr>
<tr>
<td>Prep-ved</td>
<td>proposition ved</td>
<td>ved</td>
</tr>
<tr>
<td>Pron-Acc</td>
<td>accusative pronoun</td>
<td>han, henne</td>
</tr>
<tr>
<td>Pron-Nom</td>
<td>nominative pronoun</td>
<td>han, hun</td>
</tr>
<tr>
<td>Pron-Poss-Pl</td>
<td>possessive pronoun with plural agreement</td>
<td>sine</td>
</tr>
<tr>
<td>Pron-Poss-Sg</td>
<td>possessive pronoun with singular agreement</td>
<td>sin</td>
</tr>
<tr>
<td>Prop</td>
<td>proper name</td>
<td>Oslo, Anne</td>
</tr>
<tr>
<td>Prop-Gen</td>
<td>genitive proper name</td>
<td>Alveres</td>
</tr>
<tr>
<td>Punct</td>
<td>miscellaneous punctuation</td>
<td>: [</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>comma</td>
<td>,</td>
</tr>
<tr>
<td>Punct-Quote</td>
<td>quotation marks</td>
<td>‘ ’ ’ ’</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>sentence boundary punctuation</td>
<td>… ? : !</td>
</tr>
<tr>
<td>V-Impr</td>
<td>imperative verb</td>
<td>se, ta</td>
</tr>
<tr>
<td>V-Inf</td>
<td>infinitive verb</td>
<td>kom, me, gjøre</td>
</tr>
<tr>
<td>V-Inf-SForm</td>
<td>S-Form of infinitive verb</td>
<td>kan, belastes</td>
</tr>
<tr>
<td>V-ParPart</td>
<td>past participle verb</td>
<td>reist, utpekt, slanset</td>
</tr>
<tr>
<td>V-ParPart-SForm</td>
<td>S-Form of past participle verb</td>
<td>trudser</td>
</tr>
<tr>
<td>V-Past</td>
<td>past tense verb</td>
<td>sa, vokste</td>
</tr>
<tr>
<td>V-Past-SForm</td>
<td>S-Form of past tense verb</td>
<td>levdes, instruktes</td>
</tr>
<tr>
<td>V-Pras</td>
<td>present tense verb</td>
<td>vet, gir</td>
</tr>
<tr>
<td>V-Pras-SForm</td>
<td>S-Form of present tense verb</td>
<td>flytes, mates, fiannes, synes</td>
</tr>
<tr>
<td>V-PrPar</td>
<td>present participle verb</td>
<td>voro, antalende</td>
</tr>
</tbody>
</table>
# Tagging in Nynorsk

**Figure I–38  Tagging in Nynorsk (Abbr to Aux-Pres)**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbreviation</td>
<td>abbreviation</td>
<td>red, mindre</td>
</tr>
<tr>
<td>Comparative Adjective</td>
<td>comparative adjective</td>
<td>mindre, eldre</td>
</tr>
<tr>
<td>Definite and/or Plural Adjective</td>
<td>definite and/or plural adjective</td>
<td>greite, marke, kalde</td>
</tr>
<tr>
<td>Genitive Definite Plural Adjective</td>
<td>genitive definite plural adjective</td>
<td>framandes</td>
</tr>
<tr>
<td>Definite Singular Adjective</td>
<td>definite singular adjective</td>
<td>leie, veide</td>
</tr>
<tr>
<td>Indefinite Singular Adjective</td>
<td>indefinite singular adjective</td>
<td>norsk, marke</td>
</tr>
<tr>
<td>Indefinite Indefinite Singular Adjective</td>
<td>indefinite indefinite singular adjective</td>
<td>marke, veide</td>
</tr>
<tr>
<td>Definite and/or Plural Adjective Derived from Past Participle</td>
<td>definite and/or plural adjective derived from past participle</td>
<td>kjende</td>
</tr>
<tr>
<td>Indefinite Singular Adjective Derived from Past Participle</td>
<td>indefinite singular adjective derived from past participle</td>
<td>reins, loike</td>
</tr>
<tr>
<td>Plural Adjective</td>
<td>plural adjective</td>
<td>er, isme</td>
</tr>
<tr>
<td>Adjective Derived from Present Participle</td>
<td>adjective derived from present participle</td>
<td>begynande, strande</td>
</tr>
<tr>
<td>Genitive Adjective Derived from Present Participle</td>
<td>genitive adjective derived from present participle</td>
<td>reins, loike</td>
</tr>
<tr>
<td>Superlative Adjective</td>
<td>superlative adjective</td>
<td>best, rokkost, storst</td>
</tr>
<tr>
<td>Definite Superlative Adjective</td>
<td>definite superlative adjective</td>
<td>binnen, beste, viktigaste</td>
</tr>
<tr>
<td>Adverb</td>
<td>adverb</td>
<td>leie, no</td>
</tr>
<tr>
<td>Comparative Adverb</td>
<td>comparative adverb</td>
<td>lengre, heller</td>
</tr>
<tr>
<td>Superlative Adverb</td>
<td>superlative adverb</td>
<td>verst</td>
</tr>
<tr>
<td>Imperative Auxiliary or Main Verb</td>
<td>imperative auxiliary or main verb</td>
<td>ver</td>
</tr>
<tr>
<td>Infinitive Auxiliary or Main Verb</td>
<td>infinitive auxiliary or main verb</td>
<td>bli</td>
</tr>
<tr>
<td>S-Form of Infinitive Auxiliary or Main Verb</td>
<td>S-Form of infinitive auxiliary or main verb</td>
<td>havost, faast</td>
</tr>
<tr>
<td>Past Participle of Auxiliary or Main Verb</td>
<td>past participle of auxiliary or main verb</td>
<td>hatt, vore, bli, fari</td>
</tr>
<tr>
<td>Past Tense Auxiliary or Main Verb</td>
<td>past tense auxiliary or main verb</td>
<td>hadd, var, blei, fekk</td>
</tr>
<tr>
<td>Present Tense Auxiliary or Main Verb</td>
<td>present tense auxiliary or main verb</td>
<td>har, er, blir, far</td>
</tr>
<tr>
<td>Present Participle of Auxiliary or Main Verb</td>
<td>present participle of auxiliary or main verb</td>
<td>havande, blivande, verande</td>
</tr>
<tr>
<td>Infinite Auxiliary Verb</td>
<td>infinitive auxiliary verb</td>
<td>kunne, måtte</td>
</tr>
<tr>
<td>Past Participle of Auxiliary Verb</td>
<td>past participle of auxiliary verb</td>
<td>vilja, riltta</td>
</tr>
<tr>
<td>Past Tense Auxiliary Verb</td>
<td>past tense auxiliary verb</td>
<td>kunne, måtte</td>
</tr>
<tr>
<td>Present Tense Auxiliary Verb</td>
<td>present tense auxiliary verb</td>
<td>kan, vil</td>
</tr>
</tbody>
</table>
### Figure I–39 Tagging in Nynorsk (Aux-PrPart to Prep-fra)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aux-PrPart</td>
<td>present participle of auxiliary verb</td>
<td>villande, kumande</td>
</tr>
<tr>
<td>Comp-Part</td>
<td>left compound part</td>
<td>kontor- og forretningsbygg</td>
</tr>
<tr>
<td>Coordinating</td>
<td>subordinating or relativizing conjunction</td>
<td>som, mens</td>
</tr>
<tr>
<td>Coordinating</td>
<td>subordinating or relativizing conjunction</td>
<td>og, eller</td>
</tr>
<tr>
<td>Det/PrOn-Comp</td>
<td>comparative determiner or pronoun</td>
<td>mein</td>
</tr>
<tr>
<td>Det/PrOn-Pl</td>
<td>plural determiner or pronoun</td>
<td>alle</td>
</tr>
<tr>
<td>Det/PrOn-Sg</td>
<td>singular determiner or pronoun</td>
<td>noken</td>
</tr>
<tr>
<td>Det/PrOn-Prep</td>
<td>superlative determiner or pronoun</td>
<td>mast</td>
</tr>
<tr>
<td>Det/PrOn-Prep-Def</td>
<td>definite superlative determiner or pronoun</td>
<td>mest</td>
</tr>
<tr>
<td>Det/Art-Def-Pl</td>
<td>definite plural determiner (article or</td>
<td>dei, denne</td>
</tr>
<tr>
<td>Det/Art-Def-Sg</td>
<td>definite singular determiner (article or</td>
<td>denne artikelen</td>
</tr>
<tr>
<td>Det/Art-Indef</td>
<td>indefinite singular determiner</td>
<td>eit</td>
</tr>
<tr>
<td>Interjection</td>
<td>interjection</td>
<td>hai, tjo</td>
</tr>
<tr>
<td>Def/Pl</td>
<td>definite plural noun</td>
<td>derene, armane</td>
</tr>
<tr>
<td>Def/Pl-Gen</td>
<td>genitive definite plural noun</td>
<td>salongenes, kollegenes</td>
</tr>
<tr>
<td>Def-Sg</td>
<td>definite singular noun</td>
<td>flyet, kloka</td>
</tr>
<tr>
<td>Def-Sg-Gen</td>
<td>genitive definite singular noun</td>
<td>selskapets, stovns</td>
</tr>
<tr>
<td>Def-Pl</td>
<td>indefinite plural noun</td>
<td>plassar, mappar</td>
</tr>
<tr>
<td>Def-Pl-Gen</td>
<td>genitive indefinite plural noun</td>
<td>tiders</td>
</tr>
<tr>
<td>Def-Sg</td>
<td>indefinite singular noun</td>
<td>stol, leikom</td>
</tr>
<tr>
<td>Def-Sg-Gen</td>
<td>genitive indefinite singular noun</td>
<td>fridoms</td>
</tr>
<tr>
<td>Def-Sg-Pl</td>
<td>indefinite singular or plural noun</td>
<td>lys, skritt</td>
</tr>
<tr>
<td>Letter</td>
<td>lowercase and uppercase letters</td>
<td>b, N</td>
</tr>
<tr>
<td>Net</td>
<td>URL and e-mail address</td>
<td><a href="http://www.inshift.com">www.inshift.com</a>, <a href="mailto:info@inshift.com">info@inshift.com</a></td>
</tr>
<tr>
<td>Num</td>
<td>cardinal numeric expression or plural</td>
<td>-294, 4,564, xii, 1,100</td>
</tr>
<tr>
<td>Num-Def-Sg</td>
<td>definite singular cardinal number “one”</td>
<td>to, tre</td>
</tr>
<tr>
<td>Num-Def-Sg</td>
<td>(spelled out)</td>
<td>eite</td>
</tr>
<tr>
<td>Num-Indef-Sg</td>
<td>indefinite singular cardinal number (spelled out)</td>
<td>eitt</td>
</tr>
<tr>
<td>Ord</td>
<td>ordinal number (in digits or spelled out)</td>
<td>7. , første</td>
</tr>
<tr>
<td>Part-Inf</td>
<td>infinitival particle</td>
<td>å kalle</td>
</tr>
<tr>
<td>Prep</td>
<td>preposition</td>
<td>med, ut, opp</td>
</tr>
<tr>
<td>Prep-av</td>
<td>preposition av</td>
<td>av</td>
</tr>
<tr>
<td>Prep-for</td>
<td>preposition for</td>
<td>for</td>
</tr>
<tr>
<td>Prep-frva</td>
<td>preposition frva</td>
<td>frå sin opposjon</td>
</tr>
</tbody>
</table>
### Tagging in Nynorsk (Prep-I to V-PrPart)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prep-I</td>
<td>preposition /</td>
<td>i</td>
</tr>
<tr>
<td>Prep-paa</td>
<td>preposition pa</td>
<td>paa alle</td>
</tr>
<tr>
<td>Prep-ved</td>
<td>preposition ved</td>
<td>ved</td>
</tr>
<tr>
<td>Pron-Acc</td>
<td>accusative pronoun</td>
<td>henne</td>
</tr>
<tr>
<td>Pron-Nom</td>
<td>nominative pronoun</td>
<td>hon, ho</td>
</tr>
<tr>
<td>Pron-Poss-Pl</td>
<td>possessive pronoun with plural agreement</td>
<td>sine</td>
</tr>
<tr>
<td>Pron-Poss-Sg</td>
<td>possessive pronoun with singular agreement</td>
<td>sin</td>
</tr>
<tr>
<td>Prop</td>
<td>proper name</td>
<td>Johan</td>
</tr>
<tr>
<td>Prop-Gen</td>
<td>genitive proper name</td>
<td>Egers</td>
</tr>
<tr>
<td>Punct</td>
<td>miscellaneous punctuation</td>
<td>. [ &gt;</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>comma</td>
<td></td>
</tr>
<tr>
<td>Punct-Quote</td>
<td>quotation marks</td>
<td>&quot;&quot;&quot;&quot;&quot;&quot;&quot;&quot;</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>sentence boundary punctuation</td>
<td>....? ; !</td>
</tr>
<tr>
<td>V-Impr</td>
<td>imperative verb</td>
<td>spei, kann</td>
</tr>
<tr>
<td>V-Inf</td>
<td>infinitive verb</td>
<td>gira</td>
</tr>
<tr>
<td>V-Inf-SForm</td>
<td>S-Form of infinitive verb</td>
<td>belastast, synest</td>
</tr>
<tr>
<td>V-PaPart</td>
<td>post participle verb</td>
<td>polle</td>
</tr>
<tr>
<td>V-PaPart-SForm</td>
<td>S-Form of post participle verb</td>
<td>hvitest</td>
</tr>
<tr>
<td>V-Past</td>
<td>past tense verb</td>
<td>so</td>
</tr>
<tr>
<td>V-Past-SForm</td>
<td>S-Form of past tense verb</td>
<td>mottatt, mistreiet, synest</td>
</tr>
<tr>
<td>V-Pres</td>
<td>present tense verb</td>
<td>gir, oppfattar</td>
</tr>
<tr>
<td>V-Pres-SForm</td>
<td>S-Form of present tense verb</td>
<td>first</td>
</tr>
<tr>
<td>V-PrPart</td>
<td>present participle verb</td>
<td>seg hernaande</td>
</tr>
</tbody>
</table>
### Tagging in Portuguese

**Figure I–41  Tagging in Portuguese (Adj to Det-Int)**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj</td>
<td>invariant adjective</td>
<td>simples</td>
</tr>
<tr>
<td>Adj-pl</td>
<td>plural adjective</td>
<td>colunas portuguesas</td>
</tr>
<tr>
<td>Adj-sg</td>
<td>singular adjective</td>
<td>continent europeu</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb</td>
<td>directamente</td>
</tr>
<tr>
<td>Adv-deg</td>
<td>adverbs that can modify adjectives</td>
<td>mais livre</td>
</tr>
<tr>
<td>Aux-be</td>
<td>auxiliaries of ser and estar</td>
<td>são, estão</td>
</tr>
<tr>
<td>Aux-hav</td>
<td>auxiliaries of ter and haver (&quot;have&quot;)</td>
<td>tem, háver</td>
</tr>
<tr>
<td>Aux-inf-be</td>
<td>infinitive auxiliary estar</td>
<td>estar</td>
</tr>
<tr>
<td>Aux-inf-hav</td>
<td>infinitive form of ter and haver</td>
<td>ter, haver</td>
</tr>
<tr>
<td>Aux-inf-prn-be</td>
<td>infinitive 'be' aux with attached clitic</td>
<td>sá-lo</td>
</tr>
<tr>
<td>Aux-inf-prn-have</td>
<td>int. of ter and haver with attached clitic</td>
<td>ter-se</td>
</tr>
<tr>
<td>Aux-prn-be</td>
<td>'be' auxiliaries with attached clitic</td>
<td>estava-me</td>
</tr>
<tr>
<td>Aux-prn-hav</td>
<td>auxiliaries ter and haver with attached clitic</td>
<td>tinham-me</td>
</tr>
<tr>
<td>Conj</td>
<td>unclassified conjunctions</td>
<td>nam, aquando, tal como</td>
</tr>
<tr>
<td>Conj-comp</td>
<td>comparison conjunction</td>
<td>mais do que uma vez</td>
</tr>
<tr>
<td>Conj-coord</td>
<td>coordinating conjunction</td>
<td>por fax ou correio</td>
</tr>
<tr>
<td>Conj-sub</td>
<td>subordinating conjunction</td>
<td>para que, se, que</td>
</tr>
<tr>
<td>Det/pron-dem-pl</td>
<td>plural demonstrative determiner or pronoun</td>
<td>estes, aqueles</td>
</tr>
<tr>
<td>Det/pron-dem-sg</td>
<td>singular demonstrative determiner or pronoun</td>
<td>este, aquele</td>
</tr>
<tr>
<td>Det/pron-pos-pl</td>
<td>plural possessive determiner or pronoun</td>
<td>vossos, seus</td>
</tr>
<tr>
<td>Det/pron-pos-sg</td>
<td>singular possessive determiner or pronoun</td>
<td>vossos, seu</td>
</tr>
<tr>
<td>Det/pron-quant-pl</td>
<td>plural quantifying determiner or pronoun</td>
<td>quantas vezes</td>
</tr>
<tr>
<td>Det/pron-quant-sg</td>
<td>singular quantifying determiner or pronoun</td>
<td>quanta vez</td>
</tr>
<tr>
<td>Det-int</td>
<td>interrogative determiner</td>
<td>demuestra a que ponto</td>
</tr>
</tbody>
</table>
**Figure I–42** Tagging in Portuguese (Det-Int-PI to Pron-Int-Sg)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Det-Int-PI</td>
<td>plural interrogative determiner</td>
<td>quantos, quantas, quais</td>
</tr>
<tr>
<td>Det-Int-Sg</td>
<td>singular interrogative determiner</td>
<td>quanto, quanta, qual</td>
</tr>
<tr>
<td>Det-PI</td>
<td>plural determiner</td>
<td>os maiores aplausos</td>
</tr>
<tr>
<td>Det-Rel-PI</td>
<td>plural relative determiner</td>
<td>cujos</td>
</tr>
<tr>
<td>Det-Rel-Sg</td>
<td>singular relative determiner</td>
<td>cujo</td>
</tr>
<tr>
<td>Interj</td>
<td>interjection or onomatopoeia</td>
<td>oh, claro</td>
</tr>
<tr>
<td>Num</td>
<td>numeric expression</td>
<td>123</td>
</tr>
<tr>
<td>Part-Neg</td>
<td>negation particle</td>
<td>nunca</td>
</tr>
<tr>
<td>Prep</td>
<td>preposition</td>
<td>com</td>
</tr>
<tr>
<td>Prep-a</td>
<td>preposition a</td>
<td>a</td>
</tr>
<tr>
<td>Prep-Adv</td>
<td>combination prep. and adverb</td>
<td>venho daqui</td>
</tr>
<tr>
<td>Prep-de</td>
<td>preposition de</td>
<td>de</td>
</tr>
<tr>
<td>Prep-Dem-PI</td>
<td>combination prep. and plural demonstrative</td>
<td>desses recursos</td>
</tr>
<tr>
<td>Prep-Dem-Sg</td>
<td>combination prep. and singular demonstrative</td>
<td>nesta plaza</td>
</tr>
<tr>
<td>Prep-Det-PI</td>
<td>combination prep. and det. determiner</td>
<td>nos, longe das</td>
</tr>
<tr>
<td>Prep-Det-a</td>
<td>combination a and det. determiner</td>
<td>aos</td>
</tr>
<tr>
<td>Prep-Det-Sg</td>
<td>combination a and sg. determiner</td>
<td>da, de outro</td>
</tr>
<tr>
<td>Prep-prep</td>
<td>combination prep. and pronoun</td>
<td>para</td>
</tr>
<tr>
<td>Prep-Pron</td>
<td>combination prep. and pronoun</td>
<td>atrás de</td>
</tr>
<tr>
<td>Prep-Quant-PI</td>
<td>combination prep. and plural quantifier</td>
<td>mais terrores</td>
</tr>
<tr>
<td>Prep-Quant-Sg</td>
<td>combination prep. and sg. quantifier</td>
<td>muita novem</td>
</tr>
<tr>
<td>Prep-Ael</td>
<td>combination preposition and relative pronoun</td>
<td>nesta praia onde...</td>
</tr>
<tr>
<td>Prep-Ael-Pi</td>
<td>combination preposition and plural relative pronoun</td>
<td>alunes aos quais</td>
</tr>
<tr>
<td>Prep-Ael-Sg</td>
<td>combination preposition and singular relative pronoun</td>
<td>através do qual</td>
</tr>
<tr>
<td>Pron</td>
<td>invariant pronoun</td>
<td>si</td>
</tr>
<tr>
<td>Pron-Int-PI</td>
<td>pl. interrogative pronoun</td>
<td>Qual é o livro dela?</td>
</tr>
<tr>
<td>Pron-Int-Sg</td>
<td>sg. interrogative pronoun</td>
<td>Qual é um livro dela?</td>
</tr>
<tr>
<td>Nn-Net</td>
<td>URL and e-mail address</td>
<td><a href="http://www.insight.com">www.insight.com</a>, <a href="mailto:info@insight.com">info@insight.com</a></td>
</tr>
<tr>
<td>Nn-PI</td>
<td>plural noun</td>
<td>serviços</td>
</tr>
<tr>
<td>Nn-Sg</td>
<td>singular noun</td>
<td>a estrela</td>
</tr>
<tr>
<td>Nt</td>
<td>URL and e-mail address</td>
<td><a href="http://www.insight.com">www.insight.com</a></td>
</tr>
</tbody>
</table>

**AUTO_LEXER** Parts-of-Speech Tagging  I-37
### Tagging in Portuguese (Pron-Pl to V-PrPart-Pron)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pron-PI</td>
<td>plural pronoun</td>
<td>eles</td>
</tr>
<tr>
<td>Pron-Rel</td>
<td>invariant relative pronoun</td>
<td>um ortopedista que</td>
</tr>
<tr>
<td>Pron-Rel-PI</td>
<td>plural relative pronoun</td>
<td>as instalações as quais</td>
</tr>
<tr>
<td>Pron-Rel-Sg</td>
<td>singular relative pronoun</td>
<td>o ensaio o qual</td>
</tr>
<tr>
<td>Pron-Sg</td>
<td>singular pronoun</td>
<td>olo</td>
</tr>
<tr>
<td>Prop</td>
<td>proper noun</td>
<td>Lisboa, Windows</td>
</tr>
<tr>
<td>Punct</td>
<td>other punctuation</td>
<td>: ( )</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>comma</td>
<td></td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>sentence punctuation</td>
<td>. ! ? :</td>
</tr>
<tr>
<td>V/Adj-Pa-Part</td>
<td>past participle verb or adjective</td>
<td>penetrado, referida</td>
</tr>
<tr>
<td>V-Fin</td>
<td>finite verb</td>
<td>corresponde</td>
</tr>
<tr>
<td>V-Fin-Pron</td>
<td>finite verb with attached clitic</td>
<td>deu-lhe</td>
</tr>
<tr>
<td>V-Inf</td>
<td>infinitive verb</td>
<td>reuni, conservar</td>
</tr>
<tr>
<td>V-Inf-Pron</td>
<td>infinitive verb with attached clitic</td>
<td>data-se</td>
</tr>
<tr>
<td>V-PrPart</td>
<td>present participle verb</td>
<td>falando</td>
</tr>
<tr>
<td>V-PrPart-Pron</td>
<td>present part. verb with attached clitic</td>
<td>deixando-a</td>
</tr>
</tbody>
</table>
**Tagging in Russian**

*Figure I–44  Tagging in Russian (Adj-Nom to Prep-Gen)*

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj-Nom</td>
<td>Adjective in nominative case</td>
<td>красивый, красная, красное, красные</td>
</tr>
<tr>
<td>Adj-Acc</td>
<td>Adjective in accusative case</td>
<td>красивого, красивую, красивое, красные</td>
</tr>
<tr>
<td>Adj-Gen</td>
<td>Adjective in genitive case</td>
<td>красивого, красивой, красивых</td>
</tr>
<tr>
<td>Adj-ObI</td>
<td>Adjective in oblique case</td>
<td>красивым, красивой, красивому, красивыми</td>
</tr>
<tr>
<td>Adj-Comp</td>
<td>Comparative adjective</td>
<td>чаще</td>
</tr>
<tr>
<td>Adv</td>
<td>Adverb</td>
<td>быстро</td>
</tr>
<tr>
<td>Adv-Comp</td>
<td>Comparative adverb</td>
<td>лучше</td>
</tr>
<tr>
<td>Conj</td>
<td>Conjunction</td>
<td>и, но, чтобы</td>
</tr>
<tr>
<td>Det-Nom</td>
<td>Pronominal adjective in nominative case</td>
<td>этот</td>
</tr>
<tr>
<td>Det-Acc</td>
<td>Pronominal adjective in accusative case</td>
<td>эту</td>
</tr>
<tr>
<td>Det-Gen</td>
<td>Pronominal adjective in genitive case</td>
<td>этого</td>
</tr>
<tr>
<td>Det-ObI</td>
<td>Pronominal adjective in oblique case</td>
<td>этому</td>
</tr>
<tr>
<td>Dig</td>
<td>Number (in digits)</td>
<td>1999, 1000, 05</td>
</tr>
<tr>
<td>Pron-IntrRel-Nom</td>
<td>Relative pronoun in nominative case</td>
<td>кто</td>
</tr>
<tr>
<td>Pron-IntrRel-Acc</td>
<td>Relative pronoun in accusative case</td>
<td>кого, что</td>
</tr>
<tr>
<td>Pron-IntrRel-Gen</td>
<td>Relative pronoun in genitive case</td>
<td>кого, чего</td>
</tr>
<tr>
<td>Pron-IntrRel-ObI</td>
<td>Relative pronoun in oblique case</td>
<td>кому</td>
</tr>
<tr>
<td>Interj</td>
<td>Interjection</td>
<td>да, эх, ба</td>
</tr>
<tr>
<td>Nn-Nom</td>
<td>Noun in nominative case</td>
<td>сестра, сестры</td>
</tr>
<tr>
<td>Nn-Acc</td>
<td>Noun in accusative case</td>
<td>сестру, сестер</td>
</tr>
<tr>
<td>Nn-Gen</td>
<td>Noun in genitive case</td>
<td>сестры, сестер</td>
</tr>
<tr>
<td>Nn-ObI</td>
<td>Noun in oblique case</td>
<td>сестрой, сестрашь</td>
</tr>
<tr>
<td>Num</td>
<td>Number</td>
<td>три, восемь</td>
</tr>
<tr>
<td>Ord</td>
<td>Ordinal number (in digits)</td>
<td>7, 3</td>
</tr>
<tr>
<td>Pron-Pers-Nom</td>
<td>Personal pronoun in nominative case</td>
<td>я, ты</td>
</tr>
<tr>
<td>Pron-Pers-Acc</td>
<td>Personal pronoun in accusative case</td>
<td>меня, тебя</td>
</tr>
<tr>
<td>Pron-Pers-Gen</td>
<td>Personal pronoun in genitive case</td>
<td>меня, тебя</td>
</tr>
<tr>
<td>Pron-Pers-ObI</td>
<td>Personal pronoun in oblique case</td>
<td>мной, тобой</td>
</tr>
<tr>
<td>Prep-Nom</td>
<td>Preposition governing nominative case</td>
<td>плюс, минус</td>
</tr>
<tr>
<td>Prep-Acc</td>
<td>Preposition governing accusative case</td>
<td>за</td>
</tr>
<tr>
<td>Prep-Gen</td>
<td>Preposition governing genitive case</td>
<td>без, назначу</td>
</tr>
</tbody>
</table>
### Figure I–45 Tagging in Russian (Prep-Obl to Verb-Brf)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prep-Obl</td>
<td>Preposition governing oblique case</td>
<td>благодаря, к.</td>
</tr>
<tr>
<td>Pron-Adv</td>
<td>Pronominal adverb</td>
<td>откуда, кое-как</td>
</tr>
<tr>
<td>Pron-Nom</td>
<td>Pronoun in nominative case</td>
<td>все, ничто</td>
</tr>
<tr>
<td>Pron-Acc</td>
<td>Pronoun in accusative case</td>
<td>все, ничего</td>
</tr>
<tr>
<td>Pron-Gen</td>
<td>Pronoun in genitive case</td>
<td>всего, ничего</td>
</tr>
<tr>
<td>Pron-Obl</td>
<td>Pronoun in oblique case</td>
<td>всеми, ничем</td>
</tr>
<tr>
<td>Prop-Nom</td>
<td>Proper name in nominative case</td>
<td>Москва, Мальцев</td>
</tr>
<tr>
<td>Prop-Acc</td>
<td>Proper name in accusative case</td>
<td>Москве, Мальцев</td>
</tr>
<tr>
<td>Prop-Gen</td>
<td>Proper name in genitive case</td>
<td>Москву, Мальцев</td>
</tr>
<tr>
<td>Prop-Obl</td>
<td>Proper name in oblique case</td>
<td>Москве, Мальцев</td>
</tr>
<tr>
<td>Part</td>
<td>Particle</td>
<td>аж, же</td>
</tr>
<tr>
<td>Part-Int</td>
<td>Introduction particle</td>
<td>аовсь</td>
</tr>
<tr>
<td>Part-Sent</td>
<td>Sentence particle</td>
<td>анимь</td>
</tr>
<tr>
<td>Part-Mood</td>
<td>Mood marker particle</td>
<td>бы, ли</td>
</tr>
<tr>
<td>Aux</td>
<td>Auxiliary verb</td>
<td>быть</td>
</tr>
<tr>
<td>Verb-Fin</td>
<td>Finite verb</td>
<td>делать, делят, делал</td>
</tr>
<tr>
<td>Verb-Gen</td>
<td>Adverbial participle (gerund)</td>
<td>делая, делящей, делая</td>
</tr>
<tr>
<td>Verb-Inf</td>
<td>Infinitive verb</td>
<td>делать</td>
</tr>
<tr>
<td>Verb-Acc</td>
<td>Participle in accusative case</td>
<td>делавшего, делавше, делавшую</td>
</tr>
<tr>
<td>Verb-Gen</td>
<td>Participle in genitive case</td>
<td>делавшего, делавшей</td>
</tr>
<tr>
<td>Verb-Nom</td>
<td>Participle in nominative case</td>
<td>делавший, делавшее, делавших</td>
</tr>
<tr>
<td>Verb-Obl</td>
<td>Participle in oblique case</td>
<td>делавшим, делавшей</td>
</tr>
<tr>
<td>Verb-Brf</td>
<td>Participle in brief form</td>
<td>делая, делай, деляна</td>
</tr>
</tbody>
</table>
Tagging in Slovak

**Figure I–46  Tagging in Slovak (Interj to Pron-Interrog)**

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interj</td>
<td>Interjection</td>
<td>zbohem, vyborne</td>
</tr>
<tr>
<td>Nm</td>
<td>Invariant noun</td>
<td>zombi, štipíto</td>
</tr>
<tr>
<td>Nm-Pl-Gen</td>
<td>Plural, genitive noun</td>
<td>účetov, účtovaní</td>
</tr>
<tr>
<td>Nm-Pl-Case</td>
<td>Plural, nominative, accusative, dative, locative, and instrumental noun</td>
<td>účastní, účtovaní</td>
</tr>
<tr>
<td>Nm-Sg-Gen</td>
<td>Singular, genitive noun</td>
<td>účetní, účtovaní</td>
</tr>
<tr>
<td>Nm-Sg-Case</td>
<td>Singular, nominative, accusative, dative, locative and instrumental noun</td>
<td>účetní, účtovaní</td>
</tr>
<tr>
<td>Prop</td>
<td>Proper noun</td>
<td>Swansea, Swasinami</td>
</tr>
<tr>
<td>Num</td>
<td>Number expression other than cardinal or ordinal, e.g., numbers</td>
<td>1, 12%</td>
</tr>
<tr>
<td>Num-Card</td>
<td>Cardinal number</td>
<td>osmasto, štyristie osemasto</td>
</tr>
<tr>
<td>Num-Ord</td>
<td>Ordinal number</td>
<td>dvanást, dvanáscto</td>
</tr>
<tr>
<td>Part</td>
<td>Particle</td>
<td>nie, by, na nožky</td>
</tr>
<tr>
<td>Pref</td>
<td>Stand-alone prefix</td>
<td>vodo, ne</td>
</tr>
<tr>
<td>Prep</td>
<td>Preposition</td>
<td>v, zo</td>
</tr>
<tr>
<td>Pron-Dem-Pl</td>
<td>Plural demonstrative pronoun</td>
<td>všetkalci, všetkalcom</td>
</tr>
<tr>
<td>Pron-Dem-Sg</td>
<td>Singular demonstrative pronoun</td>
<td>taký, všetkom</td>
</tr>
<tr>
<td>Pron</td>
<td>Indefinite pronoun</td>
<td>čoši</td>
</tr>
<tr>
<td>Pron-Pl</td>
<td>Plural pronoun</td>
<td>dačička, češi</td>
</tr>
<tr>
<td>Pron-Sg</td>
<td>Singular pronoun</td>
<td>kdečka, všelkačka</td>
</tr>
<tr>
<td>Pron-Interrog</td>
<td>Interrogative pronoun</td>
<td>kto, všelikoho</td>
</tr>
</tbody>
</table>

**Figure I–47  Tagging in Slovak (Punct-Quote to Punct-Close)**

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punct-Quote</td>
<td>Quote</td>
<td>&quot;</td>
</tr>
<tr>
<td>Punct</td>
<td>Other punctuation</td>
<td>...</td>
</tr>
<tr>
<td>flm-Net</td>
<td>URL, e-mail address</td>
<td><a href="http://www.inxight.com">www.inxight.com</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:info@inxight.com">info@inxight.com</a></td>
</tr>
<tr>
<td>V-Inf</td>
<td>Infinitive verb</td>
<td>účtovať, účtovat</td>
</tr>
<tr>
<td>V-Past-Pl</td>
<td>Plural, past tense verb</td>
<td>účtovali, účtkovali</td>
</tr>
<tr>
<td>V-Past-Sg</td>
<td>Singular, past tense verb</td>
<td>účtoval, účtovala</td>
</tr>
<tr>
<td>V-Prea-Pl</td>
<td>Plural, present tense verb</td>
<td>účtovája, účtujeme</td>
</tr>
<tr>
<td>V-Prea-Sg</td>
<td>Singular, present tense verb</td>
<td>účtovávam, účtujem</td>
</tr>
<tr>
<td>V-Fut-Pl</td>
<td>Plural, future tense verb</td>
<td>budú, budete</td>
</tr>
<tr>
<td>V-Fut-Sg</td>
<td>Singular, future tense verb</td>
<td>bude, budeš</td>
</tr>
<tr>
<td>V-Aux</td>
<td>Auxiliary verb</td>
<td>vš, všiš</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>Sentence ending punctuation</td>
<td>! ? .</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>Comma</td>
<td>,</td>
</tr>
<tr>
<td>Punct-Open</td>
<td>Opening punctuation</td>
<td>(</td>
</tr>
<tr>
<td>Punct-Close</td>
<td>Closing punctuation</td>
<td>)</td>
</tr>
</tbody>
</table>
Tagging in Slovenian

Figure I–48  Tagging in Slovenian (Adj to Num-Gen)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj</td>
<td>Adjectives</td>
<td>miren, mini, mlad, mladi</td>
</tr>
<tr>
<td>Adj-Comp</td>
<td>Comparative adjectives</td>
<td>lepši, lepš, bolj divji, manj divji</td>
</tr>
<tr>
<td>Adj-Sup</td>
<td>Superlative adjectives</td>
<td>najlepš, najlepše, najbolj divji, najmanj divji</td>
</tr>
<tr>
<td>Adv</td>
<td>Adverbs</td>
<td>lepo, naglo, nagloma</td>
</tr>
<tr>
<td>Adv-Comp</td>
<td>Comparative adverbs</td>
<td>lepše, bolj vroče, manj razločno</td>
</tr>
<tr>
<td>Adv-Sup</td>
<td>Superlative adverbs</td>
<td>najlepše, najbolj vroče, najmanj razločno</td>
</tr>
<tr>
<td>Conj</td>
<td>Conjunction</td>
<td>in, pa, medtem ko</td>
</tr>
<tr>
<td>Conj-Part</td>
<td>Conjunction or participle</td>
<td>samo, ne</td>
</tr>
<tr>
<td>Intj</td>
<td>Interjection</td>
<td>pfi, chej</td>
</tr>
<tr>
<td>Nn</td>
<td>Invariant noun, including abbreviation, acronym and etc.</td>
<td>ZDA, št.</td>
</tr>
<tr>
<td>Nn-Du-Gen</td>
<td>Dual, genitive noun</td>
<td>čvrstosti</td>
</tr>
<tr>
<td>Nn-Du-Case</td>
<td>Dual, nominative, accusative, dative, locative and instrumental noun</td>
<td>čvrstostih, čvrstostima</td>
</tr>
<tr>
<td>Nn-Pi-Gen</td>
<td>Plural, genitive noun</td>
<td>čvrstosti</td>
</tr>
<tr>
<td>Nn-Pi-Case</td>
<td>Plural, nominative, accusative, dative, locative and instrumental noun</td>
<td>čvrstostih, čvrstostimi</td>
</tr>
<tr>
<td>Nn-Sg-Gen</td>
<td>Singular, genitive noun</td>
<td>čvrstosti</td>
</tr>
<tr>
<td>Nn-Sg-Case</td>
<td>Singular, nominative, accusative, dative, locative and instrumental noun</td>
<td>čvrstosti, čvrststijo</td>
</tr>
<tr>
<td>Num</td>
<td>Invariant number expression</td>
<td>1, 12%</td>
</tr>
<tr>
<td>Num-Card-Gen</td>
<td>Cardinal number, genitive</td>
<td>stoš, stož</td>
</tr>
<tr>
<td>Num-Card-Case</td>
<td>Cardinal number, nominative, accusative, dative, locative and instrumental</td>
<td>sto, stožim</td>
</tr>
<tr>
<td>Num-Ord-Gen</td>
<td>Ordinal number, genitive</td>
<td>stož, stožim</td>
</tr>
<tr>
<td>Num-Ord-Case</td>
<td>Ordinal number, nominative, accusative, dative, locative and instrumental</td>
<td>sto, stožim</td>
</tr>
<tr>
<td>Num-Gen</td>
<td>Number expressions other than cardinal or ordinal numbers, genitive</td>
<td>četrtoč, četrtočik</td>
</tr>
</tbody>
</table>
### Figure I–49  Tagging in Slovenian (Num-Case to Punct-Comma)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num-Case</td>
<td>Number expressions other than cardinal or ordinal, noninative, accusative, dative, locative and instrumental</td>
<td>ṣetvocka, ṣetvorki</td>
</tr>
<tr>
<td>Part</td>
<td>Particle</td>
<td>že, zal</td>
</tr>
<tr>
<td>Prep</td>
<td>Preposition</td>
<td>pod, po</td>
</tr>
<tr>
<td>Prep-Comp</td>
<td>Preposition with clitic</td>
<td>poduje, podnjo</td>
</tr>
<tr>
<td>Pron-Dem-Du</td>
<td>Dual demonstrative pronoun</td>
<td>toliki, tolikuma</td>
</tr>
<tr>
<td>Pron-Dem-Pl</td>
<td>Plural demonstrative pronoun</td>
<td>tolikse, tolikumi</td>
</tr>
<tr>
<td>Pron-Dem-Sg</td>
<td>Singular demonstrative pronoun</td>
<td>to, toliko</td>
</tr>
<tr>
<td>Pron-Ref</td>
<td>Reflexive pronoun, invariant in number</td>
<td>sebe, seboj, sebi</td>
</tr>
<tr>
<td>Pron-Ref-Sg</td>
<td>Singular reflexive pronoun</td>
<td>svoj, svojim, svoji</td>
</tr>
<tr>
<td>Pron-Ref-Du</td>
<td>Dual reflexive pronoun</td>
<td>svoji, svojih, svojima</td>
</tr>
<tr>
<td>Pron-Ref-Pl</td>
<td>Plural reflexive pronoun</td>
<td>svoji, svoje, svoja</td>
</tr>
<tr>
<td>Pron-Pers-Sg</td>
<td>Singular personal pronoun</td>
<td>jaz, ti, on, ona, ono</td>
</tr>
<tr>
<td>Pron-Pers-Du</td>
<td>Dual personal pronoun</td>
<td>mi, me, vi, ve, oni</td>
</tr>
<tr>
<td>Pron-Pers-Pl</td>
<td>Personal pronoun, plural</td>
<td>mi, me, vi, ve, oni</td>
</tr>
<tr>
<td>Pron-Poss-Sg</td>
<td>Possessive pronoun, singular</td>
<td>svoj, svojim, svojega</td>
</tr>
<tr>
<td>Pron-Poss-Du</td>
<td>Possessive pronoun, dual</td>
<td>svoj, svojih, svojima</td>
</tr>
<tr>
<td>Pron-Poss-Pl</td>
<td>Possessive pronoun, plural</td>
<td>svoj, svoje, svoja</td>
</tr>
<tr>
<td>Pron-Interro</td>
<td>Interrogative pronoun</td>
<td>kdo, kaj, kateri</td>
</tr>
<tr>
<td>Pron-Rel</td>
<td>Relative pronoun</td>
<td>kdo, kar, kateri, ki</td>
</tr>
<tr>
<td>Pron-Pl</td>
<td>Plural pronoun</td>
<td>vsen, vesni, vse, vsa</td>
</tr>
<tr>
<td>Pron-Du</td>
<td>Dual pronoun</td>
<td>vsi, vsemi</td>
</tr>
<tr>
<td>Pron-Sg</td>
<td>Singular pronoun</td>
<td>vse, vsen, vse</td>
</tr>
<tr>
<td>Prop</td>
<td>Proper noun</td>
<td>Savva, Ljubljana</td>
</tr>
<tr>
<td>Pron</td>
<td>Other pronouns, indefinite, evaluative and etc.</td>
<td>isti, drug</td>
</tr>
<tr>
<td>V-Aux</td>
<td>Auxiliary verb</td>
<td>biti, bi</td>
</tr>
<tr>
<td>V-Sep</td>
<td>Supine verb</td>
<td>prodati, spati</td>
</tr>
<tr>
<td>V-inf</td>
<td>Infinitive verb</td>
<td>prodati, spati</td>
</tr>
<tr>
<td>V-PPast-Du</td>
<td>Dual, past tense verb</td>
<td>čevrati, čvršča</td>
</tr>
<tr>
<td>V-PPast-Pl</td>
<td>Plural, past tense verb</td>
<td>čevrati, čvršča</td>
</tr>
<tr>
<td>V-PPast-Sg</td>
<td>Singular, past tense verb</td>
<td>čevrati, čvršča</td>
</tr>
<tr>
<td>V-Pres-Du</td>
<td>Dual, present tense verb</td>
<td>jva, jis</td>
</tr>
<tr>
<td>V-Pres-Pl</td>
<td>Plural, present tense verb</td>
<td>jte, jmo</td>
</tr>
<tr>
<td>V-Pres-Sg</td>
<td>Singular, present tense verb</td>
<td>je, ješ</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>Sentence-ending punctuation</td>
<td></td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>Comma</td>
<td></td>
</tr>
</tbody>
</table>

### Figure I–50  Tagging in Slovenian (Punct-Open to Nn-Net)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punct-Open</td>
<td>Opening punctuation</td>
<td></td>
</tr>
<tr>
<td>Punct-Close</td>
<td>Closing punctuation</td>
<td></td>
</tr>
<tr>
<td>Punct-Quote</td>
<td>Quote</td>
<td></td>
</tr>
<tr>
<td>Punct</td>
<td>Other punctuation</td>
<td></td>
</tr>
<tr>
<td>Mn-Net</td>
<td>URL, e-mail address</td>
<td><a href="http://www.inight.com">www.inight.com</a>, <a href="mailto:info@inight.com">info@inight.com</a></td>
</tr>
</tbody>
</table>
**Tagging in Spanish**

*Figure I–51 Tagging in Spanish (Adj to Nn)*

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj</td>
<td>invariant adjective</td>
<td>beige, mini</td>
</tr>
<tr>
<td>Adj-Ord-Pl</td>
<td>plural, spelled out ordinal number</td>
<td>primeros, décimas</td>
</tr>
<tr>
<td>Adj-Ord-Sg</td>
<td>singular, spelled out ordinal number</td>
<td>octavo, cuarta</td>
</tr>
<tr>
<td>Adj-Pl</td>
<td>plural adjective</td>
<td>bonitos, nacionales</td>
</tr>
<tr>
<td>Adj-Sg</td>
<td>singular adjective</td>
<td>nacional</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb</td>
<td>siempre, directamente</td>
</tr>
<tr>
<td>Adv-Adj</td>
<td>adverbs that can modify adjectives</td>
<td>muy importante</td>
</tr>
<tr>
<td>Adv-Int</td>
<td>interrogative adverb</td>
<td>cuándo</td>
</tr>
<tr>
<td>Adv-Rel</td>
<td>adverbial relativizers</td>
<td>donde</td>
</tr>
<tr>
<td>Aux-be</td>
<td>the auxiliaries ser and estar ('be')</td>
<td>es, fui, estaba</td>
</tr>
<tr>
<td>Aux-have</td>
<td>the auxiliary haber ('have')</td>
<td>han, hubo, hay</td>
</tr>
<tr>
<td>Aux-inf-be</td>
<td>infinitive 'be' auxiliary</td>
<td>estar</td>
</tr>
<tr>
<td>Aux-inf-hav</td>
<td>infinitive form of haber</td>
<td>haber</td>
</tr>
<tr>
<td>Aux-inf-Pron-be</td>
<td>infinitive be auxiliary with attached clitic</td>
<td>serme, asturbe</td>
</tr>
<tr>
<td>Aux-inf-Pron-have</td>
<td>infinitive of haber with attached clitic</td>
<td>haberse, habérseme</td>
</tr>
<tr>
<td>Conj</td>
<td>conjunction</td>
<td>si, porque</td>
</tr>
<tr>
<td>Conj-coamo</td>
<td>the word 'como'</td>
<td>como</td>
</tr>
<tr>
<td>Conj-Coord</td>
<td>coordinating conjunction</td>
<td>y, o, y/b</td>
</tr>
<tr>
<td>Conj-que</td>
<td>the word 'que'</td>
<td>que</td>
</tr>
<tr>
<td>Det-Pron-Quan-Pl</td>
<td>plural quantifying determiner or pronoun</td>
<td>unas cosas</td>
</tr>
<tr>
<td>Det-Pron-Quan-Sg</td>
<td>singular quantifying determiner or pronoun</td>
<td>ningún, poca</td>
</tr>
<tr>
<td>Det-Dem-Pl</td>
<td>plural demonstrative determiner</td>
<td>estas, esos</td>
</tr>
<tr>
<td>Det-Dem-Sg</td>
<td>singular demonstrative determiner</td>
<td>esta</td>
</tr>
<tr>
<td>Det-Sg</td>
<td>singular determiner</td>
<td>tis</td>
</tr>
<tr>
<td>Det-Pre-Pl</td>
<td>plural pre-determiner</td>
<td>todas los, todos los</td>
</tr>
<tr>
<td>Det-Pre-Sg</td>
<td>singular pre-determiner</td>
<td>todo el, toda la</td>
</tr>
<tr>
<td>Det-Rel</td>
<td>relative determiner</td>
<td>cual, cierto</td>
</tr>
<tr>
<td>Det-Sg</td>
<td>singular determiner</td>
<td>mi</td>
</tr>
<tr>
<td>Interj</td>
<td>interjection or onomatopoeia</td>
<td>ah</td>
</tr>
<tr>
<td>Nn</td>
<td>noun, invariant for number</td>
<td>fénix</td>
</tr>
<tr>
<td>Tag</td>
<td>Description</td>
<td>Examples</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>Nn-Letter</td>
<td>lowercase letter with or without a period or an appendage without a period</td>
<td>a, h, M</td>
</tr>
<tr>
<td>Nn-Nit</td>
<td>URL or e-mail address</td>
<td><a href="http://www.inight.com">www.inight.com</a>, <a href="mailto:info@inight.com">info@inight.com</a></td>
</tr>
<tr>
<td>Nn-Pl</td>
<td>plural noun</td>
<td>gatos</td>
</tr>
<tr>
<td>Nn-Sg</td>
<td>singular noun</td>
<td>gato</td>
</tr>
<tr>
<td>Num</td>
<td>numeric expression, cardinal number</td>
<td>123, XIX, once, cuatrocientos</td>
</tr>
<tr>
<td>Num-Ord</td>
<td>ordinal number</td>
<td>1o., 2ª</td>
</tr>
<tr>
<td>Part-Neg</td>
<td>the negation particle no</td>
<td>no</td>
</tr>
<tr>
<td>Prep</td>
<td>preposition</td>
<td>en, con, por</td>
</tr>
<tr>
<td>Prep-a</td>
<td>preposition a</td>
<td>a casa</td>
</tr>
<tr>
<td>Prep-do</td>
<td>preposition de</td>
<td>la casa de María</td>
</tr>
<tr>
<td>Prep-De</td>
<td>combination of preposition and determiner</td>
<td>fuera del, antes del</td>
</tr>
<tr>
<td>Prep-De-a</td>
<td>combination a and determiner</td>
<td>al</td>
</tr>
<tr>
<td>Prep-De-de</td>
<td>combination de and determiner</td>
<td>del</td>
</tr>
<tr>
<td>Prep-pa</td>
<td>preposition para</td>
<td>para la casa</td>
</tr>
<tr>
<td>Pron</td>
<td>pronoun</td>
<td>yo</td>
</tr>
<tr>
<td>Pron-Clinic</td>
<td>clitic pronoun (acc. or dat.)</td>
<td>le, me, os, nos</td>
</tr>
<tr>
<td>Pron-Dem</td>
<td>demonstrative pronoun</td>
<td>ésta, aquél</td>
</tr>
<tr>
<td>Pron-Int</td>
<td>interrogative pronoun</td>
<td>cuánto, cuál, quien</td>
</tr>
<tr>
<td>Pron-Poss</td>
<td>possessive pronoun</td>
<td>el mío, las vuestras</td>
</tr>
<tr>
<td>Pron-Rel</td>
<td>relative pronoun</td>
<td>lo cual, quien</td>
</tr>
<tr>
<td>Pron-se</td>
<td>the reflexive pronoun</td>
<td>se</td>
</tr>
<tr>
<td>Prop</td>
<td>proper noun or alpha-numeric punctuation combinations</td>
<td>Pablo</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>other punctuation</td>
<td>' ' / &amp; ( ;</td>
</tr>
<tr>
<td>Punct-sent</td>
<td>sentence punctuation</td>
<td>. ! ?</td>
</tr>
<tr>
<td>V/Adv-PrPart-Pl</td>
<td>plural past participle verb or adjective</td>
<td>contenidos, hechos</td>
</tr>
<tr>
<td>V/Adv-PrPart-Sg</td>
<td>singular past participle verb or adjective</td>
<td>privado, fundido</td>
</tr>
<tr>
<td>V-Inf</td>
<td>finite verb</td>
<td>tiene, puede, dicta</td>
</tr>
<tr>
<td>V-Impr</td>
<td>imperative verb</td>
<td>deíne, sigue</td>
</tr>
<tr>
<td>V-Impr-Pron</td>
<td>imperative verb with attached clitic</td>
<td>hacerse, aguacea</td>
</tr>
<tr>
<td>V-Inf</td>
<td>infinitive verb</td>
<td>evita, tener, conducir</td>
</tr>
<tr>
<td>V-Inf-Pron</td>
<td>infinitive verb with attached clitic</td>
<td>hacerse, suprimirios</td>
</tr>
<tr>
<td>V-PrPart</td>
<td>present participle verb</td>
<td>estando, tocando</td>
</tr>
<tr>
<td>V-PrPart-Pron</td>
<td>present participle verb with attached clitic</td>
<td>haciéndose, tomando</td>
</tr>
</tbody>
</table>
Tagging in Swedish

*Figure I–53 Tagging in Swedish (Abbr to Adv)*

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbrev</td>
<td>abbreviation</td>
<td>S.</td>
</tr>
<tr>
<td>Adj-Comp</td>
<td>comparative adjective</td>
<td>äldre</td>
</tr>
<tr>
<td>Adj-Def-Pl</td>
<td>definite and/or plural adjective</td>
<td>svenska</td>
</tr>
<tr>
<td>Adj-Indef-Sg</td>
<td>indefinite singular adjective</td>
<td>grov</td>
</tr>
<tr>
<td>Adj-Sup</td>
<td>superlative adjective</td>
<td>viktigast</td>
</tr>
<tr>
<td>Adj-Sup-Def</td>
<td>definite superlative adjective</td>
<td>fullasta</td>
</tr>
<tr>
<td>Adv</td>
<td>adverb</td>
<td>redan</td>
</tr>
</tbody>
</table>
**Figure I–54  Tagging in Swedish (Cmpd-Part to Pron-Pers-Gen)**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cmpd-Part</td>
<td>compound part</td>
<td>plats- och äktomen</td>
</tr>
<tr>
<td>Conj</td>
<td>conjunction</td>
<td>att</td>
</tr>
<tr>
<td>Conj-Coord</td>
<td>coordinating conjunction</td>
<td>och, eller</td>
</tr>
<tr>
<td>Det/Pron-Pl</td>
<td>plural determiner or pronoun</td>
<td>dessa</td>
</tr>
<tr>
<td>Det/Pron-Pl-Gen</td>
<td>plural genitive determiner or pronoun</td>
<td>Anders</td>
</tr>
<tr>
<td>Det/Pron-Sg</td>
<td>singular determiner or pronoun</td>
<td>nigot, denna</td>
</tr>
<tr>
<td>Det/Pron-Sg-Gen</td>
<td>singular genitive determiner or pronoun</td>
<td>demnes</td>
</tr>
<tr>
<td>Det-Def-Pl</td>
<td>definite plural determiner</td>
<td>de partisma</td>
</tr>
<tr>
<td>Det-Def-Sg</td>
<td>definite singular determiner</td>
<td>den ekonom</td>
</tr>
<tr>
<td>Det-Indef-Sg</td>
<td>indefinite singular determiner</td>
<td>en, ett</td>
</tr>
<tr>
<td>Interj</td>
<td>conjunction</td>
<td>ja</td>
</tr>
<tr>
<td>Nn-Def-Pl</td>
<td>definite plural noun</td>
<td>verken</td>
</tr>
<tr>
<td>Nn-Def-Pl-Gen</td>
<td>genitive definite plural noun</td>
<td>professionalerna</td>
</tr>
<tr>
<td>Nn-Def-Sg</td>
<td>definite singular noun</td>
<td>historikern</td>
</tr>
<tr>
<td>Nn-Def-Sg-Gen</td>
<td>genitive definite singular noun</td>
<td>dagens</td>
</tr>
<tr>
<td>Nn-Indef-Pl</td>
<td>indefinite plural noun</td>
<td>strukturer</td>
</tr>
<tr>
<td>Nn-Indef-Pl-Gen</td>
<td>genitive indefinite plural noun</td>
<td>rikens</td>
</tr>
<tr>
<td>Nn-Indef-Sg</td>
<td>indefinite singular noun</td>
<td>dag</td>
</tr>
<tr>
<td>Nn-Indef-Sg-Gen</td>
<td>genitive indefinite singular noun</td>
<td>institutions</td>
</tr>
<tr>
<td>Nn-Indef-SP</td>
<td>indefinite singular or plural noun</td>
<td>universitet</td>
</tr>
<tr>
<td>Nn-Indef-SP-Gen</td>
<td>genitive indefinite singular or plural noun</td>
<td>års</td>
</tr>
<tr>
<td>Nn-Letter</td>
<td>lowercase and uppercase letters</td>
<td>b, N</td>
</tr>
<tr>
<td>Nn-Net</td>
<td>URL and e-mail address</td>
<td><a href="http://www.imight.com">www.imight.com</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:info@imight.com">info@imight.com</a></td>
</tr>
<tr>
<td>Num</td>
<td>cardinal number, in digits or spelled-out</td>
<td>5, 6 or 300</td>
</tr>
<tr>
<td></td>
<td>plural number</td>
<td>hundra</td>
</tr>
<tr>
<td>Num-Sg</td>
<td>spelled-out number “one”</td>
<td>en</td>
</tr>
<tr>
<td>Ord</td>
<td>ordinal number (digits or words)</td>
<td>tredje</td>
</tr>
<tr>
<td>Prep</td>
<td>preposition</td>
<td>kring</td>
</tr>
<tr>
<td>Prep-av</td>
<td>preposition av</td>
<td>av</td>
</tr>
<tr>
<td>Prep-för</td>
<td>preposition för</td>
<td>för</td>
</tr>
<tr>
<td>Prep-om</td>
<td>preposition om</td>
<td>om</td>
</tr>
<tr>
<td>Prep-på</td>
<td>preposition på</td>
<td>på</td>
</tr>
<tr>
<td>Pron-Acc</td>
<td>accusative pronoun</td>
<td>en</td>
</tr>
<tr>
<td>Pron-Gen</td>
<td>genitive pronoun</td>
<td>ens</td>
</tr>
<tr>
<td>Pron-Nom</td>
<td>nominative pronoun</td>
<td>man</td>
</tr>
<tr>
<td>Pron-Pers-Acc</td>
<td>accusative personal pronoun</td>
<td>sig</td>
</tr>
<tr>
<td>Pron-Pers-Gen</td>
<td>genitive personal pronoun</td>
<td>ders, deras</td>
</tr>
</tbody>
</table>
**Figure I–55 Tagging in Swedish (Pron-Pers-Nom to V-PrPart)**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pron-Pers-Nom</td>
<td>nominative personal pronoun</td>
<td>du</td>
</tr>
<tr>
<td>Pron-Poss-Pl</td>
<td>plural possessive pronoun</td>
<td>mina</td>
</tr>
<tr>
<td>Pron-Poss-Sg</td>
<td>singular possessive pronoun</td>
<td>vår, art</td>
</tr>
<tr>
<td>Prop</td>
<td>proper name</td>
<td>Europa, Margareta</td>
</tr>
<tr>
<td>Prop-Gen</td>
<td>genitive proper name</td>
<td>Eriks</td>
</tr>
<tr>
<td>Punct</td>
<td>punctuation</td>
<td>/,-</td>
</tr>
<tr>
<td>Punct-Comma</td>
<td>comma</td>
<td></td>
</tr>
<tr>
<td>Punct-Paren</td>
<td>bracketing punctuation</td>
<td>( ) [ ]</td>
</tr>
<tr>
<td>Punct-Quotes</td>
<td>quotation punctuation</td>
<td>‘’ ’’ &quot;&quot;</td>
</tr>
<tr>
<td>Punct-Sent</td>
<td>sentence terminating punctuation</td>
<td>, . ? !</td>
</tr>
<tr>
<td>Quant</td>
<td>quantifier</td>
<td>alla, många</td>
</tr>
<tr>
<td>V-impv</td>
<td>imperative verb</td>
<td>tänk</td>
</tr>
<tr>
<td>V-impv-SForm</td>
<td>imperative verb, S-Form</td>
<td>minas</td>
</tr>
<tr>
<td>V-inf</td>
<td>infinitive verb</td>
<td>skilla</td>
</tr>
<tr>
<td>V-inf-SForm</td>
<td>infinitive verb, S-Form</td>
<td>tänkas</td>
</tr>
<tr>
<td>V-Part</td>
<td>past participle verb</td>
<td>torknat, handiat</td>
</tr>
<tr>
<td>V-Past</td>
<td>past tense verb</td>
<td>slog</td>
</tr>
<tr>
<td>V-Past-SForm</td>
<td>past tense verb, S-Form</td>
<td>kyrktes</td>
</tr>
<tr>
<td>V-Pres</td>
<td>present tense verb</td>
<td>varnar</td>
</tr>
<tr>
<td>V-Pres-SForm</td>
<td>present tense verb, S-Form</td>
<td>sägs</td>
</tr>
<tr>
<td>V-PrPart</td>
<td>present participle verb</td>
<td>möjande</td>
</tr>
</tbody>
</table>
Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>! (operator)</td>
<td>3-47</td>
</tr>
<tr>
<td>- (operator)</td>
<td>3-30</td>
</tr>
<tr>
<td>$ (operator)</td>
<td>3-48</td>
</tr>
<tr>
<td>% (wildcard)</td>
<td>3-58</td>
</tr>
<tr>
<td>* (operator)</td>
<td>3-56</td>
</tr>
<tr>
<td>, (operator)</td>
<td>3-7</td>
</tr>
<tr>
<td>= (operator)</td>
<td>3-12, 3-13, 3-17</td>
</tr>
<tr>
<td>&gt; (operator)</td>
<td>3-51</td>
</tr>
<tr>
<td>? (operator)</td>
<td>3-18</td>
</tr>
<tr>
<td>\ (escape character)</td>
<td>4-1</td>
</tr>
<tr>
<td>_ (wildcard)</td>
<td>3-58</td>
</tr>
<tr>
<td>] (escape character)</td>
<td>4-1</td>
</tr>
</tbody>
</table>

A

ABOUT query, 3-4
- example, 3-5
  - highlight markup, 8-14, 8-26
  - highlight offsets, 8-10, 8-24
  - viewing expansion, 11-6
- accumulate operator, 3-7
- scoring, 3-7
  - stopword operator, H-3
- ADD_ATTR_SECTION procedure, 7-3
- ADD_EVENT procedure, 10-2
- ADD_FIELD_SECTION procedure, 7-4
- ADD_INDEX procedure, 7-7
- ADD_MDATA procedure, 7-9, 7-11
- ADD_MDATA_SECTION procedure, 7-12
- ADD SDATA_COLUMN procedure, 7-14
- ADD SDATA_SECTION procedure, 7-16
- ADD SPECIAL SECTION procedure, 7-18
- ADD STOP_SECTION procedure, 7-21
- ADD_STOPCLASS procedure, 7-20
- ADD_STOPTHEME procedure, 7-23
- ADD_STOPWORD procedure, 7-24
- ADD_SUB LEXER procedure, 7-26
  - example, 2-48
- ADD TRACE procedure, 10-3
- ADD ZONE SECTION procedure, 7-28
- adding a trace, 10-3
- adding an event, 10-2
- adding metadata, 7-9, 7-11, 7-12
  - adding structured data, 7-14, 7-16
- AL32UTF8 character set, 2-50, 2-51, 2-52, 2-53

ALTER INDEX
  - Add Section Constraints, 1-15
  - ALTER INDEX statement, 1-2
    - examples, 1-12
    - rebuild syntax, 1-4
    - rename syntax, 1-4
    - syntax overview, 1-2
  - ALTER TABLE
maxurls, 2-11
mixed_case, 2-44
n_table_clause, 2-76
new_german_spelling, 2-47, 16-2
newline, 2-44
no_proxy, 2-12
numgroup, 2-42
numjoin, 2-42
output_type, 2-13
override_base_letter, 16-3
p_table_clause, 2-76
path, 2-8
printjoins, 2-42
procedure, 2-13
punctuations, 2-43
r_table_clause, 2-76
setting, 7-74
skipjoins, 2-43
startjoins, 2-43
stemmer, 2-70
timeout, 2-11
urlsize, 2-11
viewing, G-7
viewing allowed values, G-7
whitespace, 2-44
AUTO stemming, 2-68
AUTO_FILTER Document Filtering Technology
  Supported Platforms for, B-2
AUTO_FILTER filter, 2-19
  and transactional CONTEXT indexes, 1-48
  character-set conversion, 2-21
  index preference object, 2-19
  setting up, B-1
  supported formats, B-4
  supported platforms, B-2
  unsupported formats, B-1
AUTO_FILTER system-defined preference, 2-85
AUTO_FILTER_OUTPUT_FORMATTING
  attribute, 2-22
AUTO_SECTION_GROUP example, 2-79
AUTO_SECTION_GROUP object, 1-56, 2-78, 7-39
AUTO_SECTION_GROUP system-defined
  preference, 2-86
automatic index synchronization, 1-7, 1-47
available traces, 10-3

B
backslash escape character, 4-1
base_letter attribute, 2-44, 16-2
base_letter_type attribute, 2-44, 16-2
base-letter conversions, 16-2
base-letter conversions, overriding, 16-3
BASIC_LEXER object, 2-39
  supported character sets, 2-39
BASIC_LEXER system-defined preference, 2-86
BASIC_LEXER type
  example, 2-47
BASIC_SECTION_GROUP object, 1-55, 2-78, 7-38
BASIC_WORDLIST object
  attributes for, 2-68
  example, 2-74
BFILE column
  indexing, 1-37
  binary attribute, 2-6, 2-15
  binary documents
    filtering, 2-4
  BINARY format column value, 1-45
BLOB column
  indexing, 1-37
  loading example, C-1
  brace escape character, 4-1
  brackets
    altering precedence, 3-3, 4-1
    grouping character, 4-1
  broader term operators
    example, 3-10
  broader term query feedback, 11-9
BROWSE_WORDS procedure, 11-2
browsing words in index, 11-2
BT function, 13-6
BT operator, 3-10
BTG function, 13-8
BTG operator, 3-10
BTI function, 13-10
BTI operator, 3-10
BTP function, 13-12
BTP operator, 3-10

C
case-sensitive
  ABOUT queries, 3-5
  case-sensitive index
    creating, 2-44
CATSEARCH operator, 1-21
CHAR column
  indexing, 1-37
character sets
  Chinese, 2-50
  Japanese, 2-51
  Korean, 2-53
characters
  continuation, 2-42
  numgroup, 2-42
  numjoin, 2-42
  printjoin, 2-42
  punctuation, 2-43
  skipjoin, 2-43
  specifying for newline, 2-44
  specifying for whitespace, 2-44
  startjoin and endjoin, 2-43
character-set
  indexing mixed columns, 2-18
  character-set conversion
with AUTO_FILTER,  2-21
charset attribute,  2-17
charset column,  1-44
CHARSET_FILTER
  attributes for,  2-17
  mixed character-set example,  2-18
Chinese
  fuzzy matching,  2-69
Chinese character sets supported,  2-50
Chinese lexicon, modifying,  15-10
Chinese text
  indexing,  2-49
CHINESE_LEXER Attribute,  2-50
CHINESE_VGRAM_LEXER object,  2-49
classifying documents,  6-2
clustering,  2-82, 6-5
CLOB column
  indexing,  1-37
clamp,  3-37
clamp size in near operator,  3-36
clustering,  2-82, 6-5
  KMEAN_CLUSTERING,  2-82
types,  2-82
CLUSTERING procedure,  6-5
clustering types,  2-82
columns types
  supported for CTXCAT index,  1-52
  supported for CTXRULE index,  1-55
  supported for CTXXPATH index,  1-56
  supported for indexing,  1-37
command attribute,  2-25
compiler, lexical,  15-10
compMem element,  2-66
composite attribute
  BASIC_LEXER,  2-45
  KOREAN_MORPH_LEXER,  2-54
composite domain index,  1-36
composite textkey
  encoding,  8-19
composite word dictionary,  2-45
composite word index
  creating for German or Dutch text,  2-45
composite words
  viewing,  11-6
concordance,  8-36
CONTAINS operator
  example,  1-33
  syntax,  1-28
CONTEXT index
  about,  1-36
  default parameters,  2-88
  syntax,  1-37
text context index type,  1-36
continuation attribute,  2-42
control file example
  SQL*Loader,  C-2
COPY_POLICY procedure,  7-31
CREATE_INDEX statement,  1-36
CONTEXT,  1-37
CTXCAT,  1-52
CTXRULE,  1-55
CTXPATH,  1-56
default parameters,  2-88
failure,  5-2
CREATE_INDEX_SCRIPT procedure,  12-5
CREATE_INDEX_SET procedure,  7-32, 7-79
CREATE_PHRASE procedure,  13-14
CREATE_POLICY procedure,  7-33
CREATE_POLICY_SCRIPT procedure,  12-6
CREATE_PREFERENCE procedure,  7-35
CREATE_RELATION procedure,  13-15
CREATE_SECTION_GROUP procedure,  7-38
CREATE_STOPLIST procedure,  7-43
CREATE_THESAURUS function,  13-17
CREATE_TRANSLATION procedure,  13-18
creating an index report,  12-3
CTX_ADM package
  MARK_FAILED,  5-2
  RECOVER,  5-3
  SET_PARAMETER,  5-4
CTX_ADM.MARK_FAILED,  5-2
CTX_ADM.CLASSES view,  G-2, G-3, G-8, G-14, G-17
CTX_CLS
  CLUSTERING,  6-5
  TRAIN,  6-2
CTX_DDL package
  ADD_ATTR_SECTION,  7-3
  ADD_FIELD_SECTION,  7-4
  ADD_MDATA,  7-9, 7-11
  ADD_MDATA_SECTION,  7-12
  ADD_SDATA_COLUMN,  7-14
  ADD_SDATA_SECTION,  7-16
  ADD_SPECIAL_SECTION,  7-18
  ADD_STOP_SECTION,  7-21
  ADD_STOPCLASS,  7-20
  ADD_STOPTHEME,  7-23
  ADD_STOPWORD,  7-24
  ADD_SUB_LEXER,  7-26
  ADD_ZONE_SECTION,  7-28
  COPY_POLICY,  7-31
  CREATE_INDEX_SET,  7-32, 7-79
  CREATE_POLICY,  7-33
  CREATE_PREFERENCE,  7-35
  CREATE_SECTION_GROUP,  7-38
  CREATE_STOPLIST,  7-43
  DROP_POLICY,  7-46
  DROP_PREFERENCE,  7-47
  DROP_STOPLIST,  7-50
  OPTIMIZE_INDEX procedure,  7-53
  REMOVE_MDATA,  7-67
  REMOVE_SECTION,  7-68
  REMOVE_STOPCLASS,  7-69
  REMOVE_STOPTHEME,  7-70
  REMOVE_STOPWORD,  7-71
  REMOVE_SUB_LEXER,  7-72
  REPLACE_INDEX_METADATA,  7-73
  SET_ATTRIBUTE,  7-74
  SYNC_INDEX procedure,  7-75
  UNSET_ATTRIBUTE,  7-78
CTX_DDL.ADD_INDEX procedure,  7-7
INDEX-IV

CTX_DOC package, 8-1
FILTER, 8-3
GIST, 8-6
HIGHLIGHT, 8-10
IFILTER, 8-13
MARKUP, 8-14
PKENCODE, 8-19
POLICY_FILTER, 8-20
POLICY_GIST, 8-21
POLICY_HIGHLIGHT, 8-24
POLICY_MARKUP, 8-28
POLICY_SNIPPET, 8-29
POLICY_THEMES, 8-31
POLICY_TOKENS, 8-33
result tables, A-5
SET_KEY_TYPE, 8-35
SNIPPET, 8-36
THEMES, 8-39
TOKENS, 8-42
CTX_DOC_KEY_TYPE system parameter, 2-88
CTX_FEEDBACK_ITEM_TYPE type, A-4
CTX_FEEDBACK_TYPE type, 11-10, A-4
CTX_INDEX_ERRORS view, G-4
example, 1-52
CTX_INDEX_OBJECTS view, G-5
CTX_INDEX_SET_INDEXES view  views
  CTX_INDEX_SET_INDEXES, G-6
  CTX_INDEX_SUB_LEXERS view, G-6, G-17
  CTX_INDEX_SUB_LEXERS_VALUES view, G-6
  CTX_INDEX_VALUES view, G-6
  CTX_INDEXES view, G-4
  CTX_OBJECT_ATTRIBUTE_LOV view, G-7
  CTX_OBJECT_ATTRIBUTES view, G-7
  CTX_OBJECTS view, G-7
CTX_OUTPUT package, 10-1
  ADD_EVENT, 10-2
  ADD_TRACE, 10-3
  DISABLE_QUERY_STATS, 10-5
  ENABLE_QUERY_STATS, 10-6
  END_LOG, 10-7
  GET_TRACE_VALUE, 10-9
  LOG_TRACES, 10-10
  LOGFILENAME, 10-11
  REMOVE_EVENT, 10-12
  REMOVE_TRACE, 10-13
  RESET_TRACE, 10-14
  START_LOG, 10-15
  CTX_PARAMETERS view, 2-87, G-8
  CTX_PENDING view, G-9
  CTX_PREFERENCE_VALUES view, G-10
  CTX_PREFERENCES view, G-10
CTX_QUERY package
  BROWSE_WORDS, 11-2
  EXPLAIN, 11-6
  HFEEDBACK, 11-9
  REMOVE_SQE, 11-13
  result tables, A-1
  STORE_SQE, 11-20
  CTX_QUERY.disable_transactional_query session
variable, 1-48
  CTX_REPORT output format, 12-3, 12-4, 12-7, 12-8, 12-17
  CTX_REPORT package, 12-1
    CREATE_INDEX_SCRIPT, 12-5
    CREATE_POLICY_SCRIPT, 12-6
    DESCRIBE_INDEX, 12-3
    DESCRIBE_POLICY, 12-4
    function versions of procedures, 12-1
    INDEX_SIZE, 12-7
    INDEX_STATS, 12-8
    QUERY_LOG_SUMMARY, 12-12
    TOKEN_INFO, 12-16
    TOKEN_TYPE, 12-18
  CTX_SECTION_GROUPS view, G-11
  CTX_SECTIONS view, G-10
  CTX_SQES view, G-11
  CTX_STOPLISTS view, G-11
  CTX_STOPWORDS view, G-11
  CTX_SUB_LEXERS view, G-12
  CTX_THES package, 13-1
    ALTER_PHRASE, 13-3
    ALTER_THESAURUS, 13-5
    BT, 13-6
    BTG, 13-8
    BTI, 13-10
    BTP, 13-12
    CREATE_PHRASE, 13-14
    CREATE_RELATION, 13-15
    CREATE_THESAURUS, 13-17
    DROP_PHRASE, 13-19
    DROP_RELATION, 13-20
    DROP_THESAURUS, 13-22
    NT, 13-25
    NTG, 13-27
    NTI, 13-29
    NTP, 13-31
    OUTPUT_STYLE, 13-33
    PT, 13-34
    result tables, A-7
    RT, 13-36
    SN, 13-38
    SYN, 13-39
    THES_TT, 13-41
    TR, 13-42
    TRSYN, 13-44
    TT, 13-46
  CTX_THESAURI view, G-12
  CTX_THES.CREATE_TRANSLATION, 13-18
  CTX_THES.DROP_TRANSLATION, 13-23
  CTX_THES.UPDATE_TRANSLATION, 13-48
  CTX_TRACE_VALUES view, G-12
  CTX_ULEXER package, 14-1
  CTX_USER_INDEX_ERRORS view, G-15
  example, 1-52
  CTX_USER_INDEX_OBJECTS view, G-15
  CTX_USER_INDEX_SET_INDEXES view, G-16
  CTX_USER_INDEX_SETS view, G-16
  CTX_USER_INDEX_SUB_LEXERS view, G-16
  CTX_USER_INDEX_VALUES view, G-17

Index-iv
Index-v

CTX_USER_INDEXES view, G-14
CTX_USER_PENDING view, G-18
CTX_USER_PREFERENCE_VALUES view, G-18
CTX_USER_PREFERENCES view, G-18
CTX_USER_SECTION_GROUPS view, G-19
CTX_USER_SECTIONS view, G-18
CTX_USER_SQES view, G-19
CTX_USER_STOPLISTS view, G-19
CTX_USER_STOPWORDS view, G-19
CTX_USER_SUB_lexERS view, G-20
CTX_USER_THES_PHRASES view, G-20
CTX_USER_THESAURUS view, G-20
CTX_VERSION view, G-20

CTXCAT index
about, 1-36
default parameters, 2-89
supported preferences, 1-53
syntax, 1-52
unsupported preferences, 1-54
tctxkbtc compiler, 15-6
tctxlc (lexical compiler), 15-10
tctxload, 15-1
examples, 15-3
import file structure, C-3

CTXRULE index
about, 1-36
and Asian languages, D-2
and multilingual support, D-2
and USER_LEXER, 2-56
default parameters, 2-90
lexer types, 1-55
syntax, 1-55

CTXRULE Index Limitations, 2-90
CTXSYS.AUTO_FILTER system preference, 2-20
CTXSYS.INSO_FILTER system preference (deprecated), 2-20

CTXXPATH index
about, 1-36
syntax, 1-56

CTXXPATH indextype
creating, 1-57

danish
alternate spelling, 16-4
index defaults, 2-85
supplied stoplist, E-3
data storage
defined procedurally, 2-12
direct, 2-3
text, 2-8
master/detail, 2-6
URL, 2-10
datastore types, 2-2
DATE column, 1-37
DBMS_PCLUTIL
BUILD_PART_INDEX, 1-51
default index
example, 1-48
default parameters
changing, 2-91
CONTEXT index, 2-88
CTXCAT index, 2-89
CTXRULE index, 2-90
viewing, 2-90
DEFAULT thesaurus, 3-10, 3-32
DEFAULT_CTXCAT_INDEX_SET system parameter, 2-89
DEFAULT_CTXCAT_LEXER system parameter, 2-89
DEFAULT_CTXCAT_STOPLIST system parameter, 2-89
DEFAULT_CTXCAT_STORAGE system parameter, 2-89
DEFAULT_CTXCAT_WORDLIST system parameter, 2-89
DEFAULT_CTXRULE_LEXER system parameter, 2-90
DEFAULT_CTXRULE_STOPLIST system parameter, 2-90
DEFAULT_CTXRULE_WORDLIST system parameter, 2-90
DEFAULT_DATASTORE system parameter, 2-88
DEFAULT_DATASTORE system-defined indexing preference, 2-85
DEFAULT_FILTER_BINARY system parameter, 2-88
DEFAULT_FILTER_FILE system parameter, 2-88
DEFAULT_FILTER_TEXT system parameter, 2-88
DEFAULT_INDEX_MEMORY system parameter, 2-87
DEFAULT_LEXER system parameter, 2-89
DEFAULT_LEXER system-defined indexing preference, 2-85
DEFAULT_RULE_STORAGE system parameter, 2-90
DEFAULT_SECTION_STORAGE system parameter, 2-89
DEFAULT_SECTION_TEXT system parameter, 2-89
DEFAULT_STOPLIST system parameter, 2-89
DEFAULT_STOPLIST system-defined preference, 2-87
DEFAULT_STORAGE system parameter, 2-89
DEFAULT_STORAGE system-defined preference, 2-87
DEFAULT_WORDLIST system parameter, 2-89
DEFAULT_WORDLIST system-defined preference, 2-87
defaults for indexing
viewing, G-8
definemerge operator, 3-12
definescore operator, 3-13
derivational stemming

enabling for English, 2-70
DESCRIBE_INDEX procedure, 12-3
DESCRIBE_POLICY procedure, 12-4
describing an index, 12-3
DETAIL_DATASTORE object, 2-6
example, 2-7
detail_key attribute, 2-6
detail_lineno attribute, 2-6
detail_table attribute, 2-6
detail_text attribute, 2-6
dictionary
  Chinese, 15-10
  Japanese, 15-10
  Korean, 2-53
  modifying, 15-10
  user, 2-45
DIRECT_DATASTORE object, 2-3
  example, 2-3
DISABLE_QUERY_STATS procedure, 10-5
disabling transactional queries, 1-48
disambiguators
  in thesaural queries, 3-10
  in thesaurus import file, C-6
DML
  affect on scoring, F-2
DML errors
  viewing, G-4
DML processing
  batch, 1-4
DML queue
  viewing, G-9
document
  classifying, 6-2
  clustering, 6-5
  filtering to HTML and plain text, 8-3
document filtering
  AUTO_FILTER, B-1
document formats
  supported, B-4
  unsupported, B-1
document loading
  SQL*Loader, C-1
document presentation
  procedures, 8-1
document services
  logging
    requests, 10-15
domain index, 1-36
double-truncated queries, 3-58
double-truncated searching
  improving performance, 2-71
DROP INDEX statement, 1-58
DROP_PHRASE procedure, 13-19
DROP_POLICY procedure, 7-46
DROP_PREFERENCE procedure, 7-47
DROP_RELATION procedure, 13-20
DROP_STOPLIST procedure, 7-50
DROP_THESAURUS procedure, 13-22
DROP_TRANSLATION procedure, 13-23
duplicating indexes with scripts, 12-5
duplicating policy with script, 12-6
Dutch
  composite word indexing, 2-45
  fuzzy matching, 2-69
  index defaults, 2-85
  supplied stoplist, E-3
E
  e-mail
    filtering and indexing, 2-21
  embedded graphics, B-9
  empty index
    POPULATE | NOPOPULATE and, 1-46
  empty indexes
    creating, 1-46
  EMPTY_STOPLIST system-defined preference, 2-87
  ENABLE_QUERY_STATS procedure, 10-6
  enabling tracing, 10-3
  END_LOG procedure, 10-7
  END_QUERY_LOG procedure, 10-8
  ending a log, 10-7
  ending a query log, 10-8
  endjoins attribute, 2-43
  English
    fuzzy matching, 2-69
    index defaults, 2-85
    supplied stoplist, E-1
  english attribute (Korean lexer), 2-54
  entity extraction, 9-1
  environment variables
    setting for AUTO_FILTER filter, B-3
  equivalence operator, 3-17
  stopword transformations, H-4
  with NEAR, 3-37
errors
  indexing, 1-52
  escaping special characters, 4-1
  event
    adding, 10-2
    removing, 10-12
  EVERY parameter, 1-7, 1-47
Example
  FILTER BY and ORDER BY with creating an
  index, 1-50
  example, 1-50
  EXP_TAB table type, A-8
  expansion operator
    soundex, 3-47
    stem, 3-48
    viewing, 11-6
  EXPLAIN procedure, 11-6
  example, 11-7
  result table, A-1
  explain table
    creating, 11-7
    retrieving data example, 11-7
  structure, A-1
  extending knowledge base, 15-6
  external filters
    specifying, 2-25
F
  failed index operation
resuming, 1-9
failure of index loading, 5-2
fast filtering, 2-22
field section
defining, 7-4
limitations, 7-5
querying, 3-60
field sections
adding dynamically, 1-10
repeated, 3-62
WITHIN example, 3-61
file data storage
example, 7-35
FILE_DATASTORE object, 2-8
example, 2-10
FILE_DATASTORE system-defined preference, 2-85
filter
INSO (deprecated), 2-19
filter attribute
MULTI_COLUMN_DATASTORE, 2-4
FILTER BY, 1-38
filter formats
supported, B-4
FILTER procedure, 8-3
example, 8-4
in-memory example, 8-4
result table, A-5
filter table
structure, A-5
filter types, 2-16
filtering
fast, with AUTO_FILTER_OUTPUT_ FORMATTING attribute, 2-22
multi_column_datastore, 2-4
stored procedures, 2-27
to plain text, 8-13
to plain text and HTML, 8-3
filters
AUTO_FILTER, 2-19, B-1
character-set, 2-17
user, 2-24
Finnish
index defaults, 2-86
supplied stoplist, E-4
format column, 1-45
formatted documents
filtering, 2-19
fragmentation of index, 1-46
French
fuzzy matching, 2-69
supplied stoplist, E-5
French stemming, 2-68
ftp_proxy attribute, 2-11
fuzzy matching
automatic language detection, 2-69
example for enabling, 2-74
specifying a language, 2-70
fuzzy operator, 3-18
fuzzy_match attribute, 2-70
fuzzy_numresults attribute, 2-71
fuzzy_score attribute, 2-71

G
German
alternate spelling attribute, 2-47
alternate spelling conventions, 2-45
composite word indexing, 2-45
fuzzy matching, 2-69
index defaults, 2-86
new spelling, querying with, 2-47, 16-2
stemming, 2-68
supplied stoplist, E-6
GET_TRACE_VALUE procedure, 10-9
Gist
generating, 8-21
gist
generating, 8-6
GIST procedure
example, 8-8
result table, A-6
updated syntax, 8-6
Gist table
structure, A-6
graphics
embedded, B-9
standalone, B-9

H
hanja attribute, 2-54
HASPATH operator, 3-20
and special characters, 3-20
HFEEDBACK procedure, 11-9
example, 11-10
result table, A-3
hierarchical query feedback information
generating, 11-9
hierarchical relationships
in thesaurus import file, C-5
HIGHLIGHT procedure, 8-10
example, 8-12
result table, A-6
highlight table
element, 8-12
structure, A-6
highlighting
generating markup, 8-14, 8-26
generating offsets, 8-10, 8-24
with NEAR operator, 3-38
homographs
in broader term queries, 3-11
in queries, 3-10
in thesaurus import file, C-6
HTML
bypassing filtering, 2-20
filtering to, 8-3
generating, 8-20
generating highlight offsets for, 8-10, 8-24
highlight markup, 8-14, 8-26
highlighting example, 8-18
indexing, 1-56, 2-21, 2-78, 7-38
zone section example, 7-28
HTML_SECTION_GROUP example, 2-79
HTML_SECTION_GROUP object, 1-56, 2-78, 7-28, 7-38
with NULL_FILTER, 2-21
HTML_SECTION_GROUP system-defined preference, 2-86
http_proxy attribute, 2-11

INDEX_PROCEDURE user_lexer attribute, 2-56
INDEX_SIZE procedure, 12-7
INDEX_STATS procedure, 12-8
index_stems
attributes, 2-41
index_stems attribute, 2-46
index_text attribute, 2-47
index_themes attribute, 2-45
indexing
master/detail example, 2-8
multilingual documents, 2-48, 2-67, D-4
parallel, 1-11, 1-42
themes, 2-45
indexing types
classifier, 2-80
clustering, 2-82
datastore, 2-2
filter, 2-16
lexer, 2-30
section group, 2-77
storage, 2-75
vs. preferences, 2-2
wordlist, 2-68
indexless document services, see policy-based document services
index-organized table, 1-36
indextype context, 1-36
inflectional stemming
enabling, 2-70
INPATH operator, 3-22
and special characters, 3-20
INPUT_TYPE user_lexer attribute, 2-57
INSERT statement
loading example, C-1
INSO_FILTER (deprecated), 2-19
inverse frequency scoring, F-1
Italian
fuzzy matching, 2-69
stemming, 2-68
supplied stoplist, E-7

JA16EUC character set, 2-51, 2-52
JA16EUCTILDE character set, 2-51, 2-52
JA16EUCYEN character set, 2-51, 2-52
JA16SJIS character set, 2-51, 2-52
JA16SJISTILDE character set, 2-51, 2-52
JA16SJISYEN character set, 2-51, 2-52
Japanese
fuzzy matching, 2-69
index defaults, 2-86
indexing, 2-51
stemming, 2-68
japanese attribute (Korean lexer), 2-54
Japanese character sets supported, 2-51
Japanese EUC character set, 2-51
Japanese lexicon, modifying, 15-10
Japanese stemming, 2-68, 3-48
JAPANESE_LEXER, 2-51
 logical operators
   with NEAR, 3-37
LONG columns
   indexing, 1-37
long_word attribute, 2-54

M
mail filter configuration file, 2-23
mail filtering, see e-mail, 2-21
MAIL_FILTER object, 2-21
MAIL_FILTER_CONFIG_FILE system parameter, 2-23
maintaining index, 1-2
MARK_FAILED procedure, 5-2
MARKUP procedure, 8-14
   example, 8-17
   HTML highlight example, 8-18
   result table, A-6
mark up table
   example, 8-17
   structure, A-6
master/detail data storage, 2-6
   example, 2-7, 7-36
master/detail tables
   indexing example, 2-8
MATCH_SCORE operator, 1-61
MATCHES operator, 1-59
MAX_INDEX_MEMORY system parameter, 2-87
max_span parameter in near operator, 3-36
maxthreads attribute, 2-11
maxurls attribute, 2-11
MDATA operator, 3-28
MDATA section, 7-9, 7-11, 7-12, 7-67
memory
   for index synchronize, 1-9
   for indexing, 1-10, 1-46, 1-56, 7-75
META tag
   creating field sections for, 7-5
   creating zone section for, 7-29
metadata, 1-6, 3-28
   replacing, 7-73
METADATA keyword, 1-6
   ALTER INDEX example, 1-13
   metadata section, 7-9, 7-11, 7-12, 7-67
MINUS operator, 3-30
   stopword transformations, H-3
mixed character-set columns
   indexing, 2-18
mixed_case attribute, 2-44
mixed-format columns
   filtering, 2-19
   indexing, 2-20
   supported formats for, B-4
   modifying user dictionary, 15-10
morpheme attribute, 2-54
MULTI_COLUMN_DATASTORE Restriction, 2-4
MULTI_LEXER object
   CREATE INDEX example, 1-49
   example, 2-48

K
k_table_clause attribute, 2-76
Key Word in Context. See KWIC
KMEAN_CLUSTERING object, 2-82
knowledge base
   supported character set, 15-6
   user-defined, 15-9
knowledge base extension compiler, 15-6
KO16KSC5601 character set, 2-53
KO16MSWIN949 character set, 2-53
Korean
   fuzzy matching, 2-69
   index defaults, 2-86
   unicode character support, 2-53
   korean character sets supported, 2-53
Korean text
   indexing, 2-53
KOREAN_MORPH_LEXER, 2-53
   composite example, 2-54
   supplied dictionaries, 2-53
   Unicode support, 2-53
KOREAN_MORPH_LEXER Attributes, 2-54
KWIC (Key Word in Context), 8-36

L
language
   setting, 2-30
language column, 1-45
left-truncated searching
   improving performance, 2-71
lexer types, 2-30
   and CTXRULE index, 1-55
lexical compiler, 15-10
lexicon. See entries under dictionary
loading text
   SQL INSERT example, C-1
   SQL*Loader example, C-1
loading thesaurus, 15-1
LOB columns
   loading, C-1
local partition index
   parallelism, 1-50
local partitioned index, 1-42
LOG_DIRECTORY system parameter, 2-88, 10-11
LOG_TRACES procedure, 10-10
LOGFILENAME procedure, 10-11
logging
   ending, 10-7
   ending a log, 10-8
   getting log file name, 10-11
   index requests, 10-15
logging queries, 12-12
logging traces, 10-10

Index-ix
MULTI_LEXER type, 2-48
MULTI_STOPLIST type, 7-43
multi-language indexing, 2-48, 2-67, 7-26, D-4
multi-language stoplist, 2-48, 2-83
multi-language tables
  querying, 1-35, 2-49
multi-lexer example
  migrating from single language, 1-13

N
n_table_clause attribute, 2-76
narrower term operators
  example, 3-32
narrower term query feedback, 11-9
NEAR operator
  backward compatibility, 3-38
  highlighting, 3-38
  scoring, 3-37
  stopword transformations, H-4
  with other operators, 3-37
  with within, 3-61
nested section searching, 3-61
nested zone sections, 7-30
nested_column attribute, 2-15
NESTED_DATASTORE attribute, 2-15
NESTED_DATASTORE object, 2-14
nested_lineno attribute, 2-15
nested_text attribute, 2-15
nested_type attribute, 2-15
new_german_spelling attribute, 2-47, 16-2
newline attribute, 2-44
NEWS_SECTION_GROUP object, 2-78, 7-39
NLS_LENGTH_SEMANTICS parameter, 1-45
no_proxy attribute, 2-12
nopopulate index parameter, 1-46
nopopulate parameter, 1-46
normalization_expr attribute, 1-30
normalized word forms, 16-1
Norwegian
  index defaults, 2-86
NOT operator, 3-40
  stopword transformations, H-3
NT function, 13-25
NT operator, 3-32
NTG function, 13-27
NTG operator, 3-32
NTI function, 13-29
NTI operator, 3-32
NTP function, 13-31
NTP operator, 3-32
NULL_FILTER object, 2-21
NULL_FILTER system-defined preference, 2-85
NULL_SECTION_GROUP object, 2-78, 7-38
NULL_SECTION_GROUP system-defined
  preference, 2-86
number attribute, 2-54
NUMBER column, 1-37
numgroup attribute, 2-42
numjoin attribute, 2-42

O
object values
  viewing, G-6
objects
  viewing index, G-7
offsets for highlighting, 8-10, 8-24
on commit, 1-7, 1-47
one_char_word attribute, 2-54
ONLINE
  CREATE_INDEX and, 1-38
OPERATION column of explain table
  values, A-2
OPERATION column of hfeedback table
  values, A-4
operator
  ABOUT, 3-4
  accumulate, 3-7
  broader term, 3-10
  equivalence, 3-17
  fuzzy, 3-18
  HASPATH, 3-20
  INPATH, 3-22
  MATCH_SCORE, 1-61
  MATCHES, 1-59
  MINUS, 3-30
  narrower term, 3-32
  NEAR
  NOT, 3-40
  OR, 3-41
  preferred term, 3-42
  related term, 3-43
  SCORE, 1-62
  scoring, 3-12, 3-13
  SDATA, 3-44
  soundex, 3-47
  SQE, 3-49
  stem, 3-48
  synonym, 3-50
  threshold, 3-51
  top term, 3-55
  TRANSFORM, 1-29
  translation term, 3-52
  translation term synonym, 3-53
  weight, 3-56
  WITHIN, 3-60
operator expansion
  viewing, 11-6
operator precedence, 3-2
  examples, 3-3
  viewing, 11-6
operators, 3-1
optimization, 7-53
  strategies, 7-53
OPTIMIZE_INDEX procedure, 7-53
optimizing index, 1-9
OPTIONS column
  explain table, A-2
  hfeedback table, A-4
OR operator, 3-41
stopword transformations, H-2
ORDER BY, 1-38
Limitations with PL/SQL Packages, 1-41
original word forms, 16-1
OUTPUT_STYLE procedure, 13-33
output_type attribute, 2-13
overlapping zone sections, 7-30
override_base_letter attribute, 16-3
overriding alternate spelling, 16-3
overriding base-letter conversions, 16-3

P
p_table_clause, 2-76
PARAGRAPHD keyword, 3-62
paragraph section
defining, 7-18
querying, 3-60
parallel index creation, 1-50
parallel indexing, 1-11, 1-42
DBMS_PCLUTIL.BUILD_PART_INDEX, 1-51
example, 1-50
local partitioned index, 1-42
parameter
transactional, 1-8, 1-47
parameters
setting, 5-4
viewing system-defined, G-8
parentheses
altering precedence, 3-3, 4-1
grouping character, 4-1
partitioned index
creating local in parallel, 1-42
example, 1-49
local, 1-42
parallel creation, 1-51
rebuild example, 1-12
partitioned index creation
eample, 1-50
partitioned tables
modifying, 1-16
parts-of-speech tagging, 1-1
path attribute, 2-8
PATH_SECTION_GROUP
querying with, 3-22
PATH_SECTION_GROUP object, 2-78, 7-39
PATH_SECTION_GROUP system-defined
preference, 2-87
pending DML
viewing, G-9
performance
wildcard searches, 3-58
PKENCODE function, 8-19
plain text
bypassing filtering, 2-20
filtering to, 8-3, 8-13
highlight markup, 8-14, 8-26
indexing with NULL_FILTER, 2-21
offsets for highlighting, 8-10
policy, 8-1
create script, 12-6
duplicate with script, 12-6
report describing, 12-4
POLICY_FILTER procedure, 8-20
POLICY_GIST procedure, 8-21
POLICY_HIGHLIGHT procedure, 8-24
POLICY_MARKUP procedure, 8-26
POLICY_SNIPPET procedure, 8-29
POLICY_THEMES procedure
syntax, 8-31
POLICY_TOKENS procedure
syntax, 8-33
policy-based document services, 8-1
populate index parameter, 1-46
populate parameter, 1-46
Portuguese
supplied stoplist, E-7
precedence of operators, 3-2
altering, 3-3, 4-1
equivalence operator, 3-17
example, 3-3
viewing, 11-6
preference classes
viewing, G-2, G-3, G-8, G-14, G-17
preference values
viewing, G-10
preferences
about, 2-1
changing, 1-6
creating, 7-35
dropping, 7-47
replacing, 1-4
specifying for indexing, 1-43
system-defined, 2-84
viewing, G-10
vs. types, 2-2
preferred term operator
eample, 3-42
prefix_index attribute, 2-71
prefix_max_length attribute, 2-72
prefix_min_length attribute, 2-72
printjoins attribute, 2-42
privileges
required for indexing, 1-37
procedure
COPY_POLICY, 7-31
CTX_DDL.ADD_INDEX, 7-7
CTX_DDL.REPLACE_INDEX_METADATA, 7-73
CTX_OUTPUT_LOG_TRACES, 10-10
CTX_OUTPUT.ADD_TRACE, 10-3
CTX_OUTPUT.END_QUERY_LOG, 10-8
CTX_OUTPUT.GET_TRACE_VALUE, 10-9
CTX_OUTPUT.REMOVE_TRACE, 10-13
CTX_OUTPUT.RESET_TRACE, 10-14
procedure attribute, 2-13
PROCEDURE_FILTER object, 2-27
progressive relaxation template, 1-29
prove_themes attribute, 2-46
proximity operator, see NEAR operator
PT function, 13-34
section group types, 2-77, 7-38
section searching, 3-60
   nested, 3-61
sections
   adding dynamically, 1-4
   constraints for dynamic addition, 1-15
   creating attribute, 7-3
   creating field, 7-4
   creating zone, 7-28
   nested, 7-30
   overlapping, 7-30
   removing, 7-68
   repeated field, 7-6
   repeated zone, 7-29
   viewing information on, G-10
SENTENCE keyword, 3-62
sentence section
   defining, 7-18
   querying, 3-60
SET_ATTRIBUTE procedure, 7-74
SET_KEY_TYPE procedure, 8-35
SET_PARAMETER procedure, 2-87, 5-4
show size of index objects, 12-7
Simplified Chinese
   index defaults, 2-86
single-byte languages
   indexing, 2-39
skipjoins attribute, 2-43
SN procedure, 13-38
SNIPPET procedure, 8-36
soundex operator, 3-47
Spanish
   fuzzy matching, 2-69
   stemming, 2-68
   supplied stoplist, E-7
special characters
   INPATH and HASPATH operators, 3-20
special section
   defining, 7-18
   querying, 3-60
spelling
   alternate, 16-2
   base letter, 16-2
   new German, 16-2
   overriding alternate, 16-3
spelling, alternate, 16-1
spelling, new German, 2-47
SQE operator, 3-49
SQL commands
   ALTER INDEX, 1-2
   CREATE INDEX, 1-36
   DROP INDEX, 1-58
SQL operators
   CONTAINS, 1-28
   MATCH_SCORE, 1-61
   MATCHES, 1-59
   SCORE, 1-62
SQL*Loader
   example, C-1
   example data file, C-2
sqlldr example, C-2
standalone graphics, B-9
START_LOG procedure, 10-15
startjoins attribute, 2-43
statistics, disabling, 10-5
statistics, enabling, 10-6
statistics, showing index, 12-8
stem indexing, 2-46
stem operator, 3-48
stemmer attribute, 2-70
stemming, 2-68, 2-70, 3-48
   automatic, 2-68
   example for enabling, 2-74
stop section
   adding dynamically, 1-11
   dynamically adding example, 1-14
stop sections
   adding, 7-21
stop_dic attribute, 2-54
stopclass
   defining, 7-20
   removing, 7-69
stoplist
   creating, 7-43
   Danish, E-3
dropping, 7-50
Dutch, E-3
English, E-1
Finnish, E-4
French, E-5
German, E-6
Italian, E-7
modifying, 2-84
multi-language, 2-48, 2-83
Portuguese, E-7
Spanish, E-7
Swedish, E-8
stoplists
   about, 2-83
   creating, 2-84
   viewing, G-11
stoptheme
   defining, 7-23
   removing, 7-70
stopword
   adding dynamically, 1-4, 1-10
defining, 7-24
   removing, 7-71
   viewing all in stoplist, G-11
stopword transformation, H-1
   viewing, 11-6
stopwords
   adding dynamically, 2-84
   removing, 2-84
storage defaults, 2-77
storage index preference
   example, 7-36
storage objects, 2-75
STORE_SQE procedure
example, 3-49
syntax, 11-20
stored queries, 3-49
stored query expression
creating, 11-20
removing, 11-13
viewing, G-19
viewing definition, G-11
structured data, 3-44
structured data section, 7-14, 7-16
sub-lexer preference
removing, 7-72
sub-lexer values
viewing, G-6
sub-lexers
viewing, G-6, G-12, G-16, G-17
substring index
example for creating, 2-74
substring_index attribute, 2-71
supplied stoplists, E-1
Supported Platforms for AUTO_FILTER, B-2
Swedish
alternate spelling, 16-4
index defaults, 2-86
supplied stoplist, E-8
SYN function, 13-39
SYN operator, 3-50
SYNC EVERY parameter, 1-7, 1-47
SYNC ON COMMIT parameter, 1-7, 1-47
sync parameter, 1-7, 1-47
SYNC_INDEX procedure, 7-75
synchronize index, 1-7, 1-47
synonym operator, 3-50
system parameters, 2-87
defaults for indexing, 2-88
system recovery
manual, 5-3
system-defined preferences, 2-84
CTXSYS.AUTO_FILTER, 2-20

T

table structure
explain, A-1
filter, A-5
Gist, A-6
hfeedback, A-3
highlight, A-6
markup, A-6
theme, A-7
tagged text
searching, 3-60
tagging for languages used with AUTO_LEXER, I-1
template query, 1-23, 1-28
text column
supported types, 1-37
Text data dictionary
cleaning up, 5-2, 5-3
TEXT format column value, 1-45
text-only index

enabling, 2-47
example, 7-35
theme functionality
supported languages, 15-9
theme highlighting
generating markup, 8-14
generating offsets, 8-10, 8-24
HTML markup example, 8-18
HTML offset example, 8-12
theme index
as default in English, 2-85
creating, 2-45
theme proving
enabling, 2-46
theme summary
generating, 8-6
generating top n, 8-8
theme table
structure, A-7
theme_language attribute, 2-46
themes
generating for document, 8-31, 8-39
generating highlight markup, 8-14, 8-26
highlight offset example, 8-12
indexing, 2-45
obtaining top n, 8-41
THEMES procedure
result table, A-7
syntax, 8-39
THES_TT procedure, 13-41
thesaurus
compiling, 15-6
creating, 13-17
creating relationships, 13-14
DEFAULT, 3-10
dropping, 13-22
import/export examples, 15-3
importing/exporting, 15-1
procedures for browsing, 13-1
renaming and truncating, 13-5
viewing information about, G-12
thesaurus import file
examples, C-7
structure, C-3
thesaurus phrases
altering, 13-3
dropping, 13-19
thesaurus relations
creating, 13-15
dropping, 13-20
thesaurus scope note
finding, 13-38
thesaurus top terms
finding all, 13-41
threshold operator, 3-51
stopword transformations, H-5
timeout attribute, 2-11
to_upper attribute, 2-54
token index optimization, 1-9
token report, generating, 12-16

Index-xiv
token, translating name into, 12-18
TOKEN_INFO procedure, 12-16
TOKEN_TYPE procedure, 12-18
TOKENS procedure
result table, A-7
syntax, 8-42
top term, 3-55
top term operator, 3-55
TR function, 13-42
TR operator, 3-52
trace value
getting, 10-9
traces, available, 10-3
 tracing
adding a trace, 10-3
available traces, 10-3
CTX_TRACE_VALUES view, G-12
enabling, 10-3
getting trace values, 10-9, G-12
logging traces, 10-10
removing trace, 10-13
resetting trace, 10-14
TRAIN procedure, 6-2
transactional CONTEXT index, 1-8, 1-47
transactional index, 11-5
transactional parameter, 1-8, 1-47
transactional text query, 1-8, 1-47
disabling, 1-48
TRANSFORM operator, 1-29
transformation
stopword, H-1
translation term operator, 3-52
translation term synonym operator, 3-53
translations
adding to thesaurus, 13-18
dropping, 13-23
English name to numeric token, 12-18
updating, 13-48
TRSYN function, 13-44
TRSYN operator, 3-53
TT function, 13-46
TT operator, 3-55
type
MULTI_LEXER, 2-48
WORLD_LEXER, 2-67, D-4
types, 2-2
indexing, 2-2
see also indexing types
U
unicode support in Korean lexer, 2-53
UNSET_ATTRIBUTE procedure, 7-78
unsupervised classification, see clustering
UPDATE GLOBAL INDEXES, 1-17, 1-18
UPDATE_TRANSLATION procedure, 13-48
URL syntax, 2-10
URL_DATASTORE object
attributes for, 2-10
dropping, 12-18
example, 2-12
URL_DATASTORE system-defined preference, 2-85
urlsize attribute, 2-11
user dictionary, modifying, 15-10
USER_DATASTORE object, 2-12
dropping, 12-18
example, 2-13
USER_DATASTORE
filtering binary documents, 8-13
user_dic attribute, 2-54
USER_FILTER object, 2-24
dropping, 2-26
USER_LEXER object, 2-55
and CTXRULE index, 2-56
UTF-16 endian auto-detection, 2-18
UTF8, 2-51
UTF8 character set, 2-39, 2-50, 2-51, 2-52, 2-53
utilities
cxtload, 15-1
V
VARCHAR2 column
indexing, 1-37
verb_adjective attribute, 2-54
version numbers
viewing, G-20
viewing
operator expansion, 11-6
operator precedence, 11-6
views
CTX_CLASSES, G-2, G-3, G-8, G-14, G-17
CTX_INDEX_ERRORS, G-4
CTX_INDEX_OBJECTS, G-5
CTX_INDEX_SUB_LEXER, G-6
CTX_INDEX_SUB_LEXERS, G-6
CTX_INDEX_SUB_LEXERS_VALUES, G-6
CTX_INDEX_VALUES, G-6
CTX_INDEXES, G-4
CTX_OBJECT_ATTRIBUTE_LOV, G-7
CTX_OBJECT_ATTRIBUTES, G-7
CTX_OBJECTS, G-7
CTX_PARAMETERS, G-8
CTX_PENDING, G-9
CTX_PREFERENCE_VALUES, G-10
CTX_PREFERENCES, G-10
CTX_SECTION_GROUPS, G-11
CTX_SECTIONS, G-10
CTX_SQES, G-11
CTX_STOPLISTS, G-11
CTX_STOPWORDS, G-11
CTX_SUB_LEXERS, G-12
CTX_THESAURI, G-12
CTX_TRACE_VALUES, G-12
CTX_USER_INDEX_ERRORS, G-15
CTX_USER_INDEXES, G-15
CTX_USER_INDEX_SET_INDEXES, G-16
CTX_USER_INDEX_SETS, G-16
CTX_USER_INDEX_SUB_LEXERS, G-16
CTX_USER_INDEX_SUB_LEXERS_VALUES, G-16
CTX_USER_INDEX_VALUES, G-17
CTX_USER_INDEXES, G-14
CTX_USER_PENDING, G-18
CTX_USER_PREFERENCE_VALUES, G-18
CTX_USER_PREFERENCES, G-18
CTX_USER_SECTION_GROUPS, G-19
CTX_USER_SECTIONS, G-18
CTX_USER_SQES, G-19
CTX_USER_STOPLISTS, G-19
CTX_USER_STOPWORDS, G-19
CTX_USER_SUB_LEXERS, G-20
CTX_USER_THES_PHRASES, G-20
CTX_USER_THESAURI, G-20
CTX_VERSION, G-20
visible flag for field sections, 7-5
visible flag in field sections, 3-61

W
weight operator, 3-56
  stopword transformations, H-5
whitespace attribute, 2-44
wildcard queries
  improving performance, 2-71
wildcard searches, 3-58
  improving performance, 3-58
wildcard_maxterms attribute, 2-72
WILDCARD_TAB type, 14-1
WITHIN operator, 3-60
  attribute sections, 3-62
  limitations, 3-60
  nested, 3-61
  precedence, 3-3
  stopword transformations, H-5
word forms, 16-1
  original vs. normalized, 16-1
WORLD_LEXER type, 2-67, D-4

X
XML documents
  attribute sections, 7-3
  doctype sensitive sections, 7-29
  indexing, 1-56, 2-78, 7-39
  querying, 3-62
XML report output format, 12-3, 12-4, 12-7, 12-8, 12-17
XML sectioning, 2-79
XML_SECTION_GROUP
  example, 2-79
XML_SECTION_GROUP object, 1-56, 2-78, 7-38
XMLType column
  indexing, 1-57

Z
ZHS16CGB231280 character set, 2-50
ZHS16GBK character set, 2-50
ZHS32GB18030 character set, 2-50
ZHT16BIG5 character set, 2-50
ZHT16HKSCS character set, 2-50
ZHT16MSWIN950 character set, 2-50
ZHT32EUC character set, 2-50
ZHT32TRIS character set, 2-50

zone section
  adding dynamically, 1-10
  creating, 7-28
  dynamically adding example, 1-14
  querying, 3-60
  repeating, 7-29