

Oracle® Outside In Search Export

Developer's Guide

Release 8.4.1

E12887-05

May 2013

Oracle Outside In Search Export Developer's Guide, Release 8.4.1

E12887-05

Copyright © 2013 Oracle and/or its affiliates. All rights reserved.

Primary Author: Mike Manier

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, the following notice is applicable:

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation shall be subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License (December 2007). Oracle America, Inc., 500 Oracle Parkway, Redwood City, CA 94065.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

This software or hardware and documentation may provide access to or information on content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services.

Contents

Preface	ix
Audience	ix
Documentation Accessibility	ix
Related Documents	ix
Conventions	ix
1 Introduction	
1.1 What's New in Release 8.4.1	1-1
1.2 What Does This Technology Do?	1-2
1.2.1 SearchML	1-2
1.2.2 PageML	1-2
1.2.3 SearchHTML	1-3
1.2.4 SearchText	1-3
1.3 Architectural Overview	1-3
1.4 Definition of Terms	1-4
1.5 Directory Structure	1-5
1.5.1 Installing Multiple SDKs	1-5
1.6 How to Use Search Export	1-5
1.7 Copyright Information	1-6
2 Windows Implementation Details	
2.1 Installation	2-1
2.1.1 NSF Support	2-2
2.2 Libraries and Structure	2-2
2.2.1 API DLLs	2-2
2.2.2 Support DLLs	2-3
2.2.3 Engine Libraries	2-4
2.2.4 Filter and Export Filter Libraries	2-4
2.2.5 Premier Graphics Filters	2-5
2.2.6 Additional Files	2-5
2.3 The Basics	2-6
2.3.1 What You Need in Your Source Code	2-6
2.3.2 Options and Information Storage	2-6
2.3.3 Structure Alignment	2-7
2.3.4 Character Sets	2-7

2.3.5	Runtime Considerations	2-7
2.4	Default Font Aliases	2-8
2.5	Changing Resources	2-9

3 UNIX Implementation Details

3.1	Installation	3-1
3.1.1	NSF Support	3-2
3.2	Libraries and Structure.....	3-2
3.2.1	API Libraries.....	3-2
3.2.2	Support Libraries	3-3
3.2.3	Engine Libraries	3-4
3.2.4	Filter and Export Filter Libraries	3-4
3.2.5	Premier Graphics Filters	3-5
3.2.6	Additional Files.....	3-5
3.3	The Basics	3-6
3.3.1	What You Need in Your Source Code	3-6
3.3.2	Information Storage.....	3-6
3.4	Character Sets	3-7
3.5	Runtime Considerations	3-7
3.5.1	X Server Requirement	3-7
3.5.2	OLE2 Objects	3-8
3.5.3	Signal Handling	3-8
3.5.4	Runtime Search Path and \$ORIGIN.....	3-8
3.6	Environment Variables	3-9
3.7	Changing Resources	3-9
3.8	HP-UX Compiling and Linking.....	3-10
3.8.1	HP-UX on RISC.....	3-10
3.8.2	HP-UX on RISC (64 bit).....	3-10
3.8.3	HP-UX on Itanium (64 bit)	3-10
3.9	IBM AIX Compiling and Linking	3-10
3.9.1	IBM AIX (32-bit pSeries)	3-11
3.9.2	IBM AIX PPC (64-bit)	3-11
3.10	Linux Compiling and Linking	3-11
3.10.1	Library Compatibility	3-11
3.10.1.1	Motif Libraries.....	3-11
3.10.1.2	GLIBC and Compiler Versions.....	3-12
3.10.1.3	Other Libraries	3-12
3.10.2	Compiling and Linking.....	3-14
3.10.2.1	Linux 32-bit, including Linux PPC.....	3-15
3.10.2.2	Linux 64-bit.....	3-15
3.10.2.3	Linux zSeries	3-15
3.11	Oracle Solaris Compiling and Linking	3-15
3.11.1	Oracle Solaris SPARC.....	3-15
3.11.2	Oracle Solaris (SPARC) 64	3-16
3.11.3	Oracle Solaris x86.....	3-16
3.12	z/OS Compiling and Linking	3-16

4 Data Access Common Functions

4.1	Deprecated Functions.....	4-2
4.2	DAInitEx.....	4-2
4.3	DADeInit.....	4-3
4.4	DAOpenDocument.....	4-3
4.4.1	IOSPECLINKEDOBJECT Structure.....	4-4
4.4.2	IOSPECARCHIVEOBJECT Structure.....	4-5
4.4.3	SCCDAOBJECT Structure.....	4-5
4.5	DACloseDocument.....	4-5
4.6	DARetrieveDocHandle.....	4-6
4.7	DASetOption.....	4-6
4.8	DAGetOption.....	4-7
4.9	DAGetFileId.....	4-7
4.10	DAGetFileIdEx.....	4-8
4.11	DAGetErrorString.....	4-9
4.12	DAGetObjectInfo.....	4-9
4.13	DAGetTreeCount.....	4-10
4.14	DAGetTreeRecord.....	4-10
4.14.1	SCCDATREENODE Structure.....	4-11
4.15	DAOpenTreeRecord.....	4-12
4.16	DAOpenRandomTreeRecord.....	4-12
4.16.1	DATREENODELOCATOR.....	4-13
4.16.2	SCCCA_TREENODELOCATOR: Tree Node Locator.....	4-13
4.17	DASaveInputObject.....	4-13
4.18	DASaveTreeRecord.....	4-14
4.19	DASaveRandomTreeRecord.....	4-15
4.19.1	DATREENODELOCATOR.....	4-16
4.19.2	SCCCA_TREENODELOCATOR: Tree Node Locator.....	4-16
4.20	DACloseTreeRecord.....	4-16
4.21	DASetStatCallback.....	4-16
4.22	DASetFileAccessCallback.....	4-17

5 Export Functions

5.1	General Functions.....	5-1
5.1.1	EXOpenExport.....	5-1
5.1.2	EXCloseExport.....	5-2
5.1.3	EXRunExport.....	5-3
5.1.4	EXExportStatus.....	5-3

6 Redirected IO

6.1	Using Redirected IO.....	6-1
6.2	Opening Files.....	6-2
6.3	IOClose.....	6-2
6.4	IORead.....	6-3
6.5	IOWrite.....	6-3
6.6	IOSeek.....	6-4

6.7	IOTell	6-5
6.8	IOGetInfo.....	6-5
6.8.1	IOGENSECONDARY and IOGENSECONDARYW Structures.....	6-8
6.8.2	File Types That Cause IOGETINFO_GENSECONDARY	6-9
6.9	IOSEEK64PROC / IOTELL64PROC	6-10
6.9.1	IOSeek64.....	6-10
6.9.2	IOTell64.....	6-10

7 Implementation Issues

7.1	Running in 24x7 Environments	7-1
7.2	Running in Multiple Threads or Processes	7-1

8 Sample Applications

8.1	Building the Samples on a Windows System	8-1
8.2	An Overview of the Sample Applications.....	8-2
8.2.1	*sample	8-2
8.2.2	batch_process_sx.....	8-2
8.2.3	export (Windows Only)	8-2
8.2.3.1	The Export Main Window.....	8-3
8.2.4	exsimple	8-4
8.2.5	extract_archive	8-4
8.2.6	process_isolate.....	8-4
8.2.7	xxredir (Search Export)	8-4
8.3	Accessing the SDK via a Java Wrapper	8-4
8.3.1	The ExJava Wrapper API.....	8-5
8.3.2	The C-Based Exporter Application	8-5
8.3.3	Compiling the Executables.....	8-6
8.3.4	The ExportTest Sample Application	8-6
8.3.5	An Example Conversion Using the ExJava Wrapper.....	8-6

A Copyrights and Licensing

A.1	Oracle Outside In Search Export Licensing	A-1
-----	---	-----

B Search Export Options

B.1	Search Export C/C++ Options.....	B-1
B.1.1	Character Mapping.....	B-1
B.1.1.1	SCCOPT_DEFAULTINPUTCHARSET	B-1
B.1.1.2	SCCOPT_UNMAPPABLECHAR.....	B-2
B.1.2	Output	B-2
B.1.2.1	SCCOPT_RENDERING_PREFER_OIT	B-2
B.1.3	Input Handling	B-3
B.1.3.1	SCCOPT_EXTRACTXMPMETADATA	B-3
B.1.3.2	SCCOPT_FALLBACKFORMAT	B-4
B.1.3.3	SCCOPT_FIFLAGS.....	B-5
B.1.3.4	SCCOPT_FORMATFLAGS.....	B-5

B.1.3.5	SCCOPT_SYSTEMFLAGS.....	B-6
B.1.3.6	SCCOPT_IGNORE_PASSWORD.....	B-6
B.1.3.7	SCCOPT_LOTUSNOTESDIRECTORY	B-6
B.1.3.8	SCCOPT_PARSEXMPMETADATA	B-7
B.1.3.9	SCCOPT_PDF_FILTER_REORDER_BIDI.....	B-7
B.1.3.10	SCCOPT_PROCESS_OLE_EMBEDDINGS	B-8
B.1.3.11	SCCOPT_TIMEZONE.....	B-9
B.1.3.12	SCCOPT_HTML_COND_COMMENT_MODE.....	B-9
B.1.3.13	SCCOPT_PDF_FILTER_DROPHYPHENS.....	B-10
B.1.3.14	SCCOPT_ARCFULLPATH	B-10
B.1.4	Compression.....	B-11
B.1.4.1	SCCOPT_FILTERLZW.....	B-11
B.1.5	XML	B-11
B.1.5.1	SCCOPT_ENABLEALLSUBOBJECTS.....	B-11
B.1.5.2	SCCOPT_XML_DEF_METHOD	B-12
B.1.5.3	SCCOPT_XML_DEF_REFERENCE	B-12
B.1.5.4	SCCOPT_XML_NULLREPLACECHAR.....	B-13
B.1.5.5	SCCOPT_XML_PAGEML_FLAGS	B-13
B.1.5.6	SCCOPT_XML_PAGEML_PRINTERNAME	B-14
B.1.5.7	SCCOPT_XML_SEARCHML_CHAR_ATTRS.....	B-14
B.1.5.8	SCCOPT_XML_SEARCHML_FLAGS.....	B-15
B.1.5.9	SCCOPT_XML_SEARCHML_OFFSET	B-17
B.1.5.10	SCCOPT_XML_SEARCHML_PARA_ATTRS.....	B-17
B.1.5.11	SCCOPT_XML_SEARCHML_UNMAPPEDTEXT	B-18
B.1.6	File System	B-19
B.1.6.1	SCCOPT_IO_BUFFER_SIZE	B-19
B.1.6.2	SCCOPT_TEMPDIR	B-20
B.1.6.3	SCCOPT_DOCUMENTMEMORYMODE	B-21
B.1.6.4	SCCOPT_REDIRECTTEMPFILE.....	B-22
B.2	Search Export SOAP Options.....	B-22
B.2.1	How Options Work	B-22
B.2.2	Character Mapping.....	B-23
B.2.2.1	defaultInputCharset	B-23
B.2.2.2	unmappableCharacter	B-23
B.2.3	Output	B-24
B.2.3.1	preferOITRendering.....	B-24
B.2.4	Input Handling	B-24
B.2.4.1	fallbackFormat	B-25
B.2.4.2	extendedTestForText.....	B-25
B.2.4.3	ignorePassword	B-26
B.2.4.4	oleEmbeddings	B-26
B.2.4.5	parseXMPMetaData	B-27
B.2.4.6	reorderBIDI.....	B-27
B.2.4.7	timezone.....	B-28
B.2.4.8	extractXMPMetaData.....	B-28
B.2.4.9	htmlCondCommentIE5On	B-28
B.2.4.10	htmlCondCommentIE6On	B-29

B.2.4.11	htmlCondCommentIE7On	B-29
B.2.4.12	htmlCondCommentIE8On	B-29
B.2.4.13	htmlCondCommentIE9On	B-29
B.2.4.14	htmlCondCommentAllOn	B-30
B.2.5	Compression	B-30
B.2.5.1	allowLZW	B-30
B.2.6	XML	B-31
B.2.6.1	allCapsOn	B-31
B.2.6.2	boldOn.....	B-31
B.2.6.3	cellInfoOn	B-31
B.2.6.4	changeNumberToTextOn.....	B-32
B.2.6.5	documentPropertiesOn.....	B-32
B.2.6.6	doubleUnderlineOn	B-32
B.2.6.7	embeddingsOn.....	B-33
B.2.6.8	errorInfoOn.....	B-33
B.2.6.9	generateSystemMetaDataOn	B-33
B.2.6.10	hiddenOn	B-34
B.2.6.11	italicOn.....	B-34
B.2.6.12	metadataOnlyOn	B-34
B.2.6.13	originalCharsetOn	B-34
B.2.6.14	outlineOn	B-35
B.2.6.15	produceURLsOn.....	B-35
B.2.6.16	revisionAddOn	B-35
B.2.6.17	revisionDeleteOn.....	B-36
B.2.6.18	revisionsOn.....	B-36
B.2.6.19	smallCapsOn.....	B-36
B.2.6.20	strikeoutOn.....	B-37
B.2.6.21	underlineOn	B-37
B.2.6.22	xmlDefinitionMethod	B-37
B.2.6.23	xmlDefinitionLocation	B-38
B.2.6.24	nullReplacementCharacter	B-38
B.2.6.25	printerName	B-38
B.2.6.26	paragraphStyleNamesOn.....	B-39
B.2.6.27	includeTextOffsets.....	B-39
B.2.6.28	paragraphAttributes.....	B-40
B.2.6.29	unmappedText.....	B-40
B.2.6.30	suppressArchiveSubDocsOn	B-41
B.2.6.31	suppressAttachmentsOn	B-41
B.2.6.32	textOutOn	B-41
B.2.6.33	xmlDeclarationOff	B-42
B.2.7	File System	B-42
B.2.7.1	fileAccess.....	B-42
B.2.7.2	readBufferSize.....	B-42
B.2.7.3	memoryMappedInputSize	B-42
B.2.7.4	tempBufferSize.....	B-43

Index

Preface

Search Export is part of Oracle's family of Original Equipment Manufacturer (OEM) technologies known as Outside In Technology, a powerful document viewing and conversion technology that can access the information in more than 600 file formats.

Audience

This document is intended for software developers who are responsible for integrating Oracle Outside In Technology into their applications.

Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc>.

Access to Oracle Support

Oracle customers have access to electronic support through My Oracle Support. For information, visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info> or visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs> if you are hearing impaired.

Related Documents

For more information, go to

<http://www.oracle.com/technetwork/indexes/documentation/index.html#middleware>

and click on Outside In Technology.

Conventions

The following text conventions are used in this document:

Convention	Meaning
boldface	Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.

Convention	Meaning
<i>italic</i>	Italic type indicates book titles, emphasis, or placeholder variables for which you supply particular values.
monospace	Monospace type indicates commands within a paragraph, URLs, code in examples, text that appears on the screen, or text that you enter.
Forward slashes (/)	Forward slashes are used to separate the directory levels in a path to a UNIX server, directory, or file. Forward slashes are also used to separate parts of an Internet address. A forward slash will always be included at the end of a UNIX directory name and might or might not be included at the end of an Internet address.
Backward slashes (\)	Backward slashes are used to separate the levels in a path to a Windows server, directory, or file. A backward slash will always be included at the end of a Windows server, directory, or file path.
<install_dir>/	This notation refers to the location on your system of the main product installation directory.

Introduction

Search Export allows developers to implement sophisticated text extraction from standard business documents. With the current version of Search Export, an application can access documents through a single C API. Search Export is ideal for a wide spectrum of applications, from rapid search and retrieval to indexing. SearchML presents the text in one of three formats: XML, HTML, or plain text.

There may be references to other Oracle Outside In Technology SDKs within this manual. To obtain complete documentation for any other Oracle Outside In product, see:

<http://www.oracle.com/technetwork/indexes/documentation/index.html#middle>
are

and click on Outside In Technology.

This chapter includes the following sections:

- [Section 1.1, "What's New in Release 8.4.1"](#)
- [Section 1.2, "What Does This Technology Do?"](#)
- [Section 1.3, "Architectural Overview"](#)
- [Section 1.4, "Definition of Terms"](#)
- [Section 1.5, "Directory Structure"](#)
- [Section 1.6, "How to Use Search Export"](#)
- [Section 1.7, "Copyright Information"](#)

1.1 What's New in Release 8.4.1

- The updated list of supported formats is linked from the page <http://www.outsideinsdk.com/>. Look for the data sheet with the latest supported formats.
- The `SCCOPT_PDF_FILTER_DROPHYPHENS` option controls whether or not the PDF filter will drop hyphens at the end of a line.
- `SCCOPT_RENDERING_PREFER_OIT` is now available on the Linux x86-64 platform.
- Work was done to avoid concatenations in text-only API's (Text Access and SearchText). The concatenations were caused when switching logical context, such as sub-documents or email fields. Previously, there were some context switches that didn't have delimiter text to separate the contexts.

- Support has been added to identify DICOM (Digital Imaging and Communications in Medicine) files.
- The following Microsoft Office formats are now supported: Microsoft Word 2011 for Mac, Microsoft Excel 2011 for Mac, Microsoft PowerPoint 2011 for Mac, Microsoft Word 2013, Microsoft Excel 2013, Microsoft PowerPoint 2013, Microsoft Outlook 2013.
- The following Adobe Creative Suite formats are now supported: Photoshop CS6, Illustrator CS6, InDesign CS6.
- Support has been added for Windows 8 and Windows 2012 Server on the Windows x86-64 platform.

1.2 What Does This Technology Do?

Search Export can normalize all of a document's content to the SearchML or PageML schemas, both provided in the form of a DTD and an XML schema, or it can output the content as simple text (the SearchText output format) or simple HTML (the SearchHTML output format). The output options available to you are determined by your license.

Note: All Search Export output formats are UTF-8 encoded Unicode text.

This section covers the following topics:

- [Section 1.2.1, "SearchML"](#)
- [Section 1.2.2, "PageML"](#)
- [Section 1.2.3, "SearchHTML"](#)
- [Section 1.2.4, "SearchText"](#)

1.2.1 SearchML

The SearchML Schema is designed to serve as a foundation for information extraction, with output that is ideal for rapid search and retrieval applications. To facilitate this purpose, the XML tags used by the SearchML schema are designed to closely mirror the information in files created by popular business applications.

Note: It is recommended that you use FI_SEARCHML_LATEST to assure that you always get the most recent SearchML schema. However, if you must have a particular version of the schema, please see `sccfi.h` for the other FI_SEARCHML* definitions.

1.2.2 PageML

The PageML output format provides information about where text would appear in a printed version of the input document. Its output consists of an XML file specifying all of the text runs for each page in the document. The text run locations are given as starting and ending character counts, or "offsets," from the beginning of the input file's text stream. This offset matches the text offsets used by Search Export's SearchML format and other members of the Oracle Outside In Viewing Technology family, including Content Access and Text Access.

The PageML Schema supports most input formats supported by Search Export. Most format types will contain <page> elements that correspond to the page that the text appears on, but there are three exceptions.

- Bitmap images have no searchable characters in the main document, so no text will appear in the output.
- All of the text for archives will appear on a single page.
- The text for spreadsheets will have each sheet appear as a separate page.

PageML is run in a manner much like other Search Export output filters, such as FI_SEARCHHTML_LATEST. When PageML formatted XML is desired, FI_PAGEML is passed as the output formatdwOutputId to EXOpenExport(). Similarly, PageML uses a new schema, also called PageML, when generating the XML output. There is a small set of options that may be used to modify its behavior:

- SCCOPT_XML_PAGEML_FLAGS
- SCCOPT_XML_PAGEML_PRINTERNAME
- textOutOn
- xmlDeclarationOff

The PageML Schema supports all word processing formats supported by Search Export, including but not limited to Microsoft Word 97 and newer, WordPerfect Version 7 and newer, HTML, ASCII, and RTF. There is also limited support for PDF.

1.2.3 SearchHTML

This format produces output that uses standard HTML tags, but will not be viewable HTML. It is a form of HTML that is easily parsed and therefore ideal for search and retrieval or indexing applications.

Document properties will be stored in <meta> tags using the name attribute for the property type and the content attribute for the property's content. The title document property will be represented by a <title> tag.

Bold, italic, and underline character attributes will be reflected using the , <i> and <u> tags respectively.

SearchHTML is run in a manner much like other Search Export output filters, such as FI_SEARCHHTML_LATEST. When SearchHTML formatted output is desired, FI_SEARCHHTML is passed as the output formatdwOutputId to EXOpenExport().

The output will obey the HTML 4.01 Transitional DTD, available at <http://www.w3.org/TR/REC-html40/>.

1.2.4 SearchText

This output format produces simple, text-only output. When extended characters are encountered, they will be output as UTF-8 encoded Unicode characters.

SearchText is run in a manner much like other Search Export output filters, such as FI_SEARCHHTML_LATEST. When SearchText formatted output is desired, FI_SEARCHTEXT is passed as the output formatdwOutputId to EXOpenExport().

1.3 Architectural Overview

The basic architecture of Oracle Outside In technologies is the same across all supported platforms:

Filter/Module	Description
Input Filter	The input filters form the base of the architecture. Each one reads a specific file format or set of related formats and sends the data to OIT through a standard set of function calls. There are more than 150 of these filters that read more than 600 distinct file formats. Filters are loaded on demand by the data access module.
Export Filter	Architecturally similar to input filters, export filters know how to write out a specific format based on information coming from the chunker module. The export filters generate XML, HTML, or text.
Chunker	The Chunker module is responsible for caching a certain amount of data from the filter and returning this data to the export filter.
Export	The Export module implements the export API and understands how to load and run individual export filters.
Data Access	The Data Access module implements a generic API for access to files. It understands how to identify and load the correct filter for all the supported file formats. The module delivers to the developer a generic handle to the requested file, which can then be used to run more specialized processes, such as the Export process.
Schema	Schemas provide a means for defining the structure, content and semantics of XML documents. Your Search Export license may include the SearchML schema. Schemas can be presented in the form of a DTD (Document Type Definition) or XML Schema (schema). The Search ML schema is provided in both forms.

1.4 Definition of Terms

The following terms are used in this documentation.

Term	Definition
Developer	Someone integrating this technology into another technology or application. Most likely this is you, the reader.
Source File	The file the developer wishes to export.
Output File	The file being written: XML, HTML, or text.
Data Access Module	The core of Oracle Outside In Data Access, in the <i>SCCDA</i> library.
Data Access Submodule (also referred to as "Submodule")	This refers to any of the Oracle Outside In Data Access modules, including <i>SCCEX</i> (Export), but excluding <i>SCCDA</i> (Data Access).
Document Handle (also referred to as "hDoc")	A Document Handle is created when a file is opened using Data Access (see Chapter 4, "Data Access Common Functions"). Each Document Handle may have any number of Subhandles.

Term	Definition
Subhandle (also referred to as "hItem")	Any of the handles created by a Submodule's <code>Open</code> function. Every Subhandle has a Document Handle associated with it. For example, the <code>hExport</code> returned by <code>EXOpenExport</code> is a Subhandle. The <code>DASetOption</code> and <code>DAGetOption</code> functions in the Data Access Module may be called with any Subhandle or Document Handle. The <code>DARetrieveDocHandle</code> function returns the Document Handle associated with any Subhandle.

1.5 Directory Structure

Each Oracle Outside In product has an `sdk` directory, under which there is a subdirectory for each platform on which the product ships (for example, `sx/sdk/sx_win-x86-32_sdk`). Under each of these directories are the following three subdirectories:

- **docs**: Contains both a PDF and HTML version of the product manual.
- **redist**: Contains only the files that the customer is allowed to redistribute. These include all the compiled modules, filter support files, `.xsd` and `.dtd` files, `cmmmap000.bin`, and third-party libraries, like `freetype`.
- **sdk**: Contains the other subdirectories that used to be at the root-level of an `sdk` (`common`, `lib` (windows only), `resource`, `samplefiles`, and `samplecode` (previously `samples`). In addition, one new subdirectory has been added, `demo`, that holds all of the compiled sample apps and other files that are needed to demo the products. These are files that the customer should not redistribute (`.cfg` files, `exportmaps`, etc.).

In the root platform directory (for example, `sx/sdk/sx_win-x86-32_sdk`), there are two files:

- **README** : Explains the contents of the `sdk`, and that `makedemo` must be run in order to use the sample applications.
- **makedemo** (either `.bat` or `.sh` – platform-based): This script will either copy (on Windows) or Symlink (on Unix) the contents of `.../redist` into `.../sdk/demo`, so that sample applications can then be run out of the `demo` directory.

1.5.1 Installing Multiple SDKs

If you load more than one OIT SDK, you must copy files from the secondary installations into the top-level OIT SDK directory as follows:

- **docs** – copy all subdirectories named “[product name]guide” into this directory.
- **redist** – copy all binaries into this directory.
- **sdk** – this directory has several subdirectories: `common`, `demo`, `lib`, `resource`, `samplecode`, `samplefiles`. In each case, copy all of the files from the secondary installation into the top-level OIT SDK subdirectory of the same name. If the top-level OIT SDK directory lacks any directories found in the directory being copied from, just copy those directories over.

1.6 How to Use Search Export

Here’s a step-by-step overview of how to export a source file.

1. Call `DAIniExt` to initialize the Data Access technology. This function needs to be called only once per application. If using threading, then pass in the correct `ThreadOption`.
2. Set any options that require a NULL handle type (optional). Certain options need to be set before the desired source file is opened. These options are identified by requiring a NULL handle type. They include, but aren't limited to:
 - `SCCOPT_FALLBACKFORMAT`
 - `SCCOPT_FIFLAGS`
 - `SCCOPT_TEMPDIR`
 - `SCCOPT_IO_BUFFERSIZE`
3. Open the Source File. `DAOpenDocument` is called to create a document handle that uniquely identifies the source file. This handle may be used in subsequent calls to the `EXOpenExport` function or the open function of any other Data Access Submodule, and will be used to close the file when access is complete. This allows the file to be accessed from multiple Data Access Submodules without reopening.
4. Set the Options. If you require option values other than the default settings, call `DASetOption` to set options. Note that options listed in the Options Guide as having "Handle Types" that accept `VTHEXPORT` may be set any time before `EXRunExport` is called. For more information on options and how to set them, see [Section 4.7, "DASetOption."](#)
5. Open a Handle to Search Export. Using the document handle, `EXOpenExport` is called to obtain an export handle that identifies the file to the specific export product. This handle will be used in all subsequent calls to the specific export functions. The `dwOutputId` parameter of this function is used to specify that the output file type should be set to one of the following:
 - `FI_SEARCHML_LATEST`
 - `FI_PAGEML`
 - `FI_SEARCHHTML`
 - `FI_SEARCHTEXT`
6. Export the File. `EXRunExport` is called to generate the output file(s) from the source file.
7. Close Handle to Search Export. `EXCloseExport` is called to terminate the export process for the file. After this function is called, the export handle will no longer be valid, but the document handle may still be used.
8. Close the Source File. `DACloseDocument` is called to close the source file. After calling this function, the document handle will no longer be valid.
9. Close Search Export. `DADeInit` is called to de-initialize the Data Access technology.

1.7 Copyright Information

The following notice must be included in the documentation, help system, or About box of any software that uses any of Oracle's executable code:

Oracle Outside In Search Export © 1991, 2013 Oracle.

The following notice must be included in the documentation of any software that uses Oracle's TIF6 filter (this filter reads TIFF and JPEG formats):

The software is based in part on the work of the Independent JPEG Group.

Windows Implementation Details

The Windows implementation of this software is delivered as a set of DLLs. For a list of the currently supported platforms, see:

<http://www.oracle.com/technetwork/indexes/documentation/index.html#middlewares>

Click on Outside In Technology, then click the Certification Information PDF.

The 64-bit version of `sccvw.dll` will not load on an AMD-64 system without Visual C++ runtime version 8 installed. This happens because the system is missing the `msvcr80.dll` library, which is required. Users can download the required library from the following location:

<http://www.microsoft.com/downloads/details.aspx?FamilyId=90548130-4468-4BBC-9673-D6ACABD5D13B&displaylang=en>

This chapter includes the following sections:

- [Section 2.1, "Installation"](#)
- [Section 2.2, "Libraries and Structure"](#)
- [Section 2.3, "The Basics"](#)
- [Section 2.4, "Default Font Aliases"](#)
- [Section 2.5, "Changing Resources"](#)

2.1 Installation

To install the demo version of the SDK, copy the contents of the ZIP archive (available on the web site) to a local directory of your choice.

This product requires the Visual C++ libraries included in the Visual C++ Redistributable Package available from Microsoft. There are three versions of this package (x86, x64, and IA64) for each corresponding version of Windows. These can be downloaded from www.microsoft.com/downloads, by searching on the site for the following packages:

- `vc redistrib_x86.exe`
- `vc redistrib_x64.exe`
- `vc redistrib_IA64.exe`

The required download version is the "2005 SP1 Redistributable Package."

Oracle Outside In requires the `msvcr80.dll` redistributable module.

The installation directory should contain the following directory structure.

Directory	Description
\docs	Includes HTML and PDF versions of the manual you are reading right now. Release notes contain more up-to-the-minute information on product changes which occurred after documentation production.
\redist	Contains a working copy of the Windows version of the technology.
\sdk\common	Contains the C include files needed to build or rebuild the technology.
\sdk\demo	Contains the compiled executables of the sample applications.
\sdk\lib	Contains the library (.lib) files needed for the products.
\sdk\resource	Contains localization resource files.
\sdk\samplecode	Contains a subdirectory holding the source code for a sample application.
sdk\samplefiles	Contains sample input files authored in a variety of popular graphics, word processor, compression, spreadsheet and presentation applications, designed to exercise Search Export.

2.1.1 NSF Support

Notes Storage Format (NSF) files are produced by the Lotus Notes Client or the Lotus Domino server. The NSF filter is the only Oracle Outside In filter that requires the native application to be present to filter the input documents. Due to integration with an outside application, NSF support will not work with redirected I/O, when an NSF file is embedded in another file, or with IOTYPE_UNICODEPATH. Either Lotus Notes version 8 or Lotus Domino version 8 must be installed on the same machine as OIT. A 32-bit version of the Lotus software must be used if you are using a 32-bit version of OIT. A 64-bit version of the Lotus software must be used if you are using a 64-bit version of OIT. On Windows, SCCOPT_LOTUSNOTESDIRECTORY should be set to the directory containing the nnotes.dll. NSF support is only available on the Win32, Win x86-64, Linux x86-32, and Solaris Sparc 32 platforms.

2.2 Libraries and Structure

The following is an overview of the files in the main installation directory for all five Oracle Outside In export products.

2.2.1 API DLLs

These libraries implement the API. They should be linked with the developer's application. Files with a .lib extension are included in the SDK.

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
sccda.dll	Data Access module	X	X	X	X	X
sccex.dll	Export module	X	X	X	X	X

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
sccfi.dll	File Identification module (identifies files based on their contents).	X	X	X	X	X

The File ID Specification may not be used directly by any application or workflow without it being separately licensed expressly for that purpose.

2.2.2 Support DLLs

The following libraries are used for support.

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
ccflex.dll	A data model adapter that converts from stream model utilized by Oracle Outside In filters to the FlexionDoc Tree model used as a basis by XML Export.					X
libexpatw.dll	A third-part XML parser					X
ocemul.dll	Output component emulation module	X	X	X	X	X
ospdf.dll	PDF generation module			X		
oswin*.dll	Interface to the native GDI implementation oswin32.dll is the 32-bit version, oswin64.dll is the 64-bit version	X	X		X	X
sccanno.dll	The annotation module	X	X	X		
sccca.dll	Content Access module (provides organized chunker data for the developer)	X	X	X		
sccch.dll	Chunker (provides caching of and access to filter data for the export engines)	X	X	X	X	X
sccdu.dll	Display Utilities module (includes text formatting)	X	X	X	X	X
sccexind.dll	The core engine for all Search Export formats: SearchText, SearchHTML, SearchML and PageML				X	
sccfmt.dll	Formatting module (resolves numbers to formatted strings)	X	X	X	X	X

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
sccfut.dll	Filter utility module	X	X	X	X	X
sccind.dll	Indexing engine. In Search Export, it handles common functionality.	X	X	X	X	
scclo.dll	Localization library (all strings, menus, dialogs and dialog procedures reside here)	X	X	X	X	X
sccole2.dll	OLE rendering module	X	X	X	X	X
sccsd.dll	Schema Definition Module Manager (brokers multiple Schema Definition Modules)					X
sccut.dll	Utility functions, including IO subsystem	X	X	X	X	X
sccxt.dll	XTree module					X
sdflex.dll	Schema Definition module (handles conversion of XML string names and attribute values to compact binary representations and vice versa)					X
wvcore.dll	The GDI Abstraction layer	X	X	X	X	X

2.2.3 Engine Libraries

The following libraries are used for display purposes.

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
debmp.dll	Raster rendering engine (TIFF, GIF, BMP, PNG, PCX...)			X		X
devect.dll	Vector/Presentation rendering engine (PowerPoint, Impress, Freelance...)	X	X	X		X
dess.dll	Spreadsheet/Database (Excel, Calc, Lotus 123...)		X	X		X
detree.dll	Archive (ZIP, GZIP, TAR...)		X	X		
dewp.dll	Document (Word, Writer, WordPerfect...)		X	X	X	

2.2.4 Filter and Export Filter Libraries

The following libraries are used for filtering.

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
vs*.dll	Filters for specific file types (there are more than 150 of these filters, covering more than 600 file formats)	X	X	X	X	X
oitnsf.id	Support file for the vsnsf filter.	X	X	X	X	X
exgdsf.dll	Export filter for GIF, JPEG, and PNG graphics files	X				X
eximg.dll	Extended image conversion module		X			
exhtml.dll	Export filter for HTML files	X				
exihtml.dll	Export filter for SearchHTML				X	
exitext.dll	Export filter for SearchText				X	
exixml.dll	Export filters for XML files using the SearchML schema				X	
expage.dll	Export filter for XML files using the PageML schema				X	
expagelayout.dll	Page layout module			X		
exxml.dll	XML Export module					X
sccimg.dll	Image conversion module	X	X			X

2.2.5 Premier Graphics Filters

The following are graphics filters.

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
i*2.flt	30 .flt files (import filters for premier graphics formats)	X	X	X	X	X
isgdi32.dll	Interface to premier graphics filters	X	X	X	X	X

2.2.6 Additional Files

The following files are also used.

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
adinit.dat	Support file for the vsacd2 filter	X	X	X	X	X
ccbf.dll	Internal				X	
cmmmap000.bin	Tables for character mapping (all character sets)	X	X	X	X	X
cmmmap000.sbc	Tables for character mapping (single-byte character sets). This file is located in the /common directory.	X	X	X	X	X

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
cmmmap000.dbc	Identical to cmmmap000.bin, but renamed for clarity (.dbc = double-byte character). This file is located in the common directory.	X	X	X	X	X
exbf.dll	Internal				X	
pageml.dtd	The Document Type Definition for the PageML schema				X	
pageml.xsd	The Extensible Schema Definition for the PageML schema				X	
searchml3.dtd	The Document Type Definitions for the SearchML schema				X	
searchml3.xsd	The Extensible Schema Definitions for the SearchML schema				X	

2.3 The Basics

The following is a discussion of some basic usage and installation features.

All the steps outlined in this section are used in the sample applications provided with the SDK. Looking at the code for the exsimple sample application is recommended for those wishing to see a real-world example of this process.

2.3.1 What You Need in Your Source Code

Any source code that uses this product should `#include` the file `sccex.h` and `#define` `WINDOWS` and `WIN32` or `WIN64`. For example, a Windows application might have a source file with the following lines:

```
#define WINDOWS          /* Will be automatically defined if your
                        compiler defines _WINDOWS */

#define WIN32
#include <sccex.h>
```

The developer's application should be linked to the product DLLs through the provided libraries.

2.3.2 Options and Information Storage

This software is based on the Oracle Outside In Viewer Technology (or simply "Viewer Technology"). When using the Export products, a list of options and a list of available display engines are built by the technology, usually the first time the product runs. You

do not need to ship these lists with your application. The lists are automatically recreated if corrupted or deleted.

The files used to store this information are stored in an .oit subdirectory in \Documents and Settings*user name*\Application Data.

If an .oit directory does not exist in the user's directory, the directory is created automatically. The files are automatically regenerated if corrupted or deleted.

The files are:

- *.d = Display Engine lists
- *.opt = Persistent options

Some applications and services may run under a local system account for which there is no users "application data" folder. The technology first does a check for an environment variable called OIT_DATA_PATH. Then it checks for APPDATA, and then LOCALAPPDATA. If none of those exist, the options files are put into the executable path of the UT module.

These file names are intended to be unique enough to avoid conflict for any combination of machine name and install directory. This allows the user to run products in separate directories without having to reload the files above. The file names are built from an 11-character string derived from the directory the Oracle Outside In technology resides in and the name of the machine it is being run on. The string is generated by code derived from the RSA Data Security, Inc. MD5 Message-Digest Algorithm.

The software still functions if these lists cannot be created for some reason. In that situation, however, significant performance degradation should be expected.

2.3.3 Structure Alignment

Oracle Outside In is built with 8-byte structure alignment. This is the default setting for most Windows compilers. This and other compiler options that should be used are demonstrated in the files provided with the sample applications in samples\win.

2.3.4 Character Sets

The strings passed in the Windows API are ANSI1252 by default.

To optimize performance on systems that do not require DBCS support, a second character mapping bin file, that does not contain any of the DBCS pages, is now included. The second bin file gives additional performance benefits for English documents, but cannot handle DBCS documents. To use the new bin file, replace the cmmmap000.bin with the new bin file, cmmmap000.sbc. For clarity, a copy of the cmmmap000.bin file (cmmmap000.dbc) is also included. Both cmmmap000.sbc and cmmmap000.dbc are located in the \common directory of the technology.

Note: All of the Search Export flavors produce most text in UTF-8 encoded Unicode. Two exceptions to this are the characters in <unmapped> elements and XMP metadata (which is passed through without character mapping being applied).

2.3.5 Runtime Considerations

The files used by the product must be in the same directory as the developer's executable.

2.4 Default Font Aliases

The technology includes the following default font alias map for Windows. The first value is the original font, the second is the alias.

- Andale Mono = Liberation Mono
- Courier = Liberation Mono
- Courier New = Liberation Mono
- Lucida Console = Liberation Mono
- Monaco = Liberation Mono
- MS Gothic = Liberation Mono
- MS Mincho = Liberation Mono
- OCR A Extended = Liberation Mono
- OCR B = Liberation Mono

- Agency FB = Liberation Sans
- Albany = Liberation Sans
- Arial = Liberation Sans
- Arial Black = Liberation Sans
- Arial Narrow = Liberation Sans
- Arial Rounded MT = Liberation Sans
- Arial Unicode MS = Liberation Sans
- Berline Sans FB = Liberation Sans
- Calibri = Liberation Sans
- Frank Gothic Demi = Liberation Sans
- Frank Gothic Medium Cond = Liberation Sans
- Franklin Gothic Book = Liberation Sans
- Futura = Liberation Sans
- Geneva = Liberation Sans
- Gill Sans = Liberation Sans
- Gill Sans MT = Liberation Sans
- Lucida Sans Regular = Liberation Sans
- Lucida Sans Unicode = Liberation Sans
- Modern No. 20 = Liberation Sans
- Tahoma = Liberation Sans
- Trebuchet MS = Liberation Sans
- Tw Cen MT = Liberation Sans
- Verdana = Liberation Sans
- Franklin Demi Cond = Liberation Sans

- Franklin Gothic = Liberation Sans
- Franklin Gothic Heavy = Liberation Sans

- Algerian = Liberation Serif
- Baskerville = Liberation Serif
- Bell MT = Liberation Serif
- Bodoni MT = Liberation Serif
- Bodoni MT Black = Liberation Serif
- Book Antiqua = Liberation Serif
- Bookman Old Style = Liberation Serif
- Calisto MT = Liberation Serif
- Cambria = Liberation Serif
- Centaur = Liberation Serif
- Century = Liberation Serif
- Century Gothic = Liberation Serif
- Century Schoolbook = Liberation Serif
- Elephant = Liberation Serif
- Footlight MT Light = Liberation Serif
- Garamond = Liberation Serif
- Georgia = Liberation Serif
- Goudy Old Style = Liberation Serif
- Lucida Bright = Liberation Serif
- MS Serif = Liberation Serif
- New York = Liberation Serif
- Palatino = Liberation Serif
- Perpetua = Liberation Serif
- Times = Liberation Serif
- times = Liberation Serif
- Times New Roman = Liberation Serif

2.5 Changing Resources

Oracle Outside In Search Export ships with the necessary files for OEMs to change any of the strings in the technology as they see fit.

Strings are stored in the lodlgstr.h file found in the resource directory. The file can be edited using any text editor.

Note: Do not directly edit the `scclo.rc` file. Strings are saved with their identifiers in `lodlgstr.h`. If a new `scclo.rc` file is saved, it will contain numeric identifiers for strings, instead of their `#define`'d names.

Once the changes have been made, the updated `scclo.dll` file can be rebuilt using the following steps:

1. Compile the `.res` file:

```
rc /fo ".\scclo.res" /i "<path to header (.h) files folder>" /d "NDEBUG"  
scclo.rc
```

2. Link the `scclo.res` file you've created with the `scclo.obj` file found in the resource directory to create a new `scclo.dll`:

```
link /DLL /OUT:scclo.dll scclo.obj scclo.res
```

Note: Developers should make sure they have set up their environment variables to build the library for their specific architecture. For Windows `x86_32`, when compiling with VS 2005, the solution is to run `vsvars32.bat` (in a standard VS 2005 installation, this is found in `C:\Program Files\Microsoft Visual Studio 8\Common7\Tools\`). If this works correctly, you will see the statement, "Setting environment for using Microsoft Visual Studio 2005 x86 tools." If you do not complete this step, you may have conflicts that lead to unresolved symbols due to conflicts with the Microsoft CRT.

3. Embed the manifest (which is created in the `\resource` directory during step 2) into the new DLL:

```
mt -manifest scclo.dll.manifest -outputresource:scclo.dll;2
```

If you are not using Microsoft Visual Studio, substitute the appropriate development tools from your environment.

Note: In previous versions of Oracle Outside In, it was possible to directly edit the `SCCLO.DLL` using Microsoft Visual Studio. Oracle Outside In DLLs are now digitally signed. Editing the signed DLL is not advisable.

UNIX Implementation Details

The UNIX implementation of the Export product set is delivered as a set of shared libraries. For a list of the currently supported platforms, see:

<http://www.oracle.com/technetwork/indexes/documentation/index.html#middleware>

Click on Outside In Technology, then click the Certification Information PDF.

This chapter includes the following sections:

- [Section 3.1, "Installation"](#)
- [Section 3.2, "Libraries and Structure"](#)
- [Section 3.3, "The Basics"](#)
- [Section 3.4, "Character Sets"](#)
- [Section 3.5, "Runtime Considerations"](#)
- [Section 3.6, "Environment Variables"](#)
- [Section 3.7, "Changing Resources"](#)
- [Section 3.8, "HP-UX Compiling and Linking"](#)
- [Section 3.9, "IBM AIX Compiling and Linking"](#)
- [Section 3.10, "Linux Compiling and Linking"](#)
- [Section 3.11, "Oracle Solaris Compiling and Linking"](#)
- [Section 3.12, "z/OS Compiling and Linking"](#)

3.1 Installation

To install the demo version of the SDK, copy the tgz file corresponding to your platform (available on the web site) to a local directory of your choice. Decompress the tgz file and then extract from the resulting tar file as follows:

```
gunzip tgzfile
tar xvf tarfile
```

The installation directory should contain the following directory structure:

Directory	Description
/docs	Includes HTML and PDF versions of the manual you are reading right now.
/redist	Contains a working copy of the UNIX version of the technology.

Directory	Description
/sdk/common	Contains the C include files needed to build or rebuild the technology.
/sdk/demo	Contains the compiled executables of the sample applications.
/sdk/resource	Contains localization resource files. For details, see Section 3.7, "Changing Resources."
/sdk/samplecode	Contains a subdirectory holding the source code for a sample application. For details, see Chapter 8, "Sample Applications."
/sdk/samplefiles	Contains sample input files authored in a variety of popular graphics, word processor, compression, spreadsheet and presentation applications, designed to exercise Search Export.

3.1.1 NSF Support

Notes Storage Format (NSF) files are produced by the Lotus Notes Client or the Lotus Domino server. The NSF filter is the only Outside In filter that requires the native application to be present to filter the input documents. Due to integration with an outside application, NSF support will not work with redirected I/O nor will it work when an NSF file is embedded in another file. Lotus Domino version 8 must be installed on the same machine as OIT. The NSF filter is currently only supported on the Win32, Win x86-64, Linux x86-32, and Solaris Sparc 32 platforms. SCCOPT_LOTUSNOTESDIRECTORY is a Windows-only option and is ignored on Unix.

Additional steps must be taken to prepare the system. It is necessary to know the name of the directory in which Lotus Domino has been installed. On Linux, this default directory is /opt/ibm/lotus/notes/latest/linux. On Solaris, it is /opt/ibm/lotus/notes/latest/sunspa.

- In the Lotus Domino directory, check for the existence of a file called "notes.ini". If the file "notes.ini" does not exist, create it in that directory and ensure that it contains the following single line:

```
[Notes]
```

- Add the Lotus Domino directory to the \$LD_LIBRARY_PATH environment variable.
- Set the environment variable \$Notes_ExecDirectory to the Lotus Domino directory.

3.2 Libraries and Structure

On UNIX platforms the Oracle Outside In products are delivered with a set of shared libraries. All libraries should be installed to a single directory. Depending upon your application, you may also need to add that directory to the system's runtime search path. For more information, see [Section 3.6, "Environment Variables."](#)

The following is a brief description of the included libraries and support files. In instances where a file extension is listed as .*, the file extension varies for each UNIX platform (**sl** on HP-UX, **so** on Linux and Solaris).

3.2.1 API Libraries

These libraries implement the API. They should be linked with the developer's application.

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
libsc_da.*	Data Access module	X	X	X	X	X

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
libsc_ex.*	Export module	X	X	X	X	X
libsc_fi.*	File Identification module (identifies files based on their contents).	X	X	X	X	X

The File ID Specification may not be used directly by any application or workflow without it being separately licensed expressly for that purpose.

3.2.2 Support Libraries

The following libraries are used for support.

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
liboc_emul.*	Output component emulation module	X	X	X	X	X
libos_gd.*	The internal rendering GDI implementation. 32-bit Linux and Solaris Sparc only.	X	X		X	X
libos_xwin.*	The native GDI implementation	X	X		X	X
libsc_anno.*	The annotation module	X	X	X		
libsc_ca.*	Content Access module (provides organized chunker data for the developer)	X	X	X		
libsc_ch.*	Chunker (provides caching of and access to filter data for the export engines)	X	X	X	X	X
libsc_du.*	Display Utilities module (includes text formatting)	X	X	X	X	X
libsc_exind.*	The core engine for all Search Export formats: SearchText, SearchHTML, SearchML and PageML				X	
libsc_fmt.*	Formatting module (resolves numbers to formatted strings)	X	X	X	X	X
libsc_fut.*	Filter utility module	X	X	X	X	X
libsc_ind.*	Indexing engine. In Search Export, it handles common functionality.	X	X	X	X	
libsc_lo.*	Localization library (all strings, menus, dialogs and dialog procedures reside here)	X	X	X	X	X
libsc_ut.*	Utility functions, including IO subsystem	X	X	X	X	X
libsc_xp.*	XPrinter bridge	X	X		X	X
libwv_core.*	The Abstraction layer	X	X	X	X	X

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
libwv_gdlib.so	The GDI rendering module. 32-bit Linux and Solaris Sparc only.	X	X		X	X

3.2.3 Engine Libraries

The following libraries are used for display purposes.

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
libde_bmp.*	Raster rendering engine (TIFF, GIF, BMP, PNG, PCX...)			X		X
libde_vect.*	Vector/Presentation rendering engine (PowerPoint, Impress, Freelance...)	X	X	X		X
libde_ss.*	Spreadsheet/Database (Excel, Calc, Lotus 123...)		X	X		X
libde_tree*	Archive (ZIP, GZIP, TAR...)		X	X		
libde_wp.*	Document (Word, Writer, WordPerfect...)		X	X	X	

3.2.4 Filter and Export Filter Libraries

The following libraries are used for filtering.

libex_gdsf must be linked with libsc_img.* at compile time. This forces the filter to be dependent on libsc_img.* at runtime, even though that module may not be used directly. If you want to reduce your application's physical footprint, you can experiment with unlinking libsc_img.*.

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
libvs_*.*	Filters for specific file types (there are more than 150 of these filters, covering more than 600 file formats)	X	X	X	X	X
libex_gdsf.*	Export filter for GIF, JPEG, and PNG graphics files	X				X
libsc_img.*	Image conversion module	X	X			X
libex_itext.*	Export filter for SearchText				X	
libex_html.*	Export filter for HTML files	X				
libex_img.*	Extended image conversion module		X			
libex_xml.*	Export filter for XML files using the Flexiondoc schema					X
libex_page.*	Export filter for XML files using the PageML schema				X	
libex_pagelayout.*	Page Layout module			X		
libex_ixml.*	Export filters for XML files using the SearchML schema				X	
libex_ihtml.*	Export filter for SearchHTML				X	

3.2.5 Premier Graphics Filters

The following are graphics filters.

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
i*2.flr	These 30 .flr files are the import filters for premier graphics formats	X	X	X	X	X
isunx2.flr	Interface to premier graphics filters	X	X	X	X	X

3.2.6 Additional Files

The following files are also used.

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
adinit.dat	Support file for the vsacad and vsacd2 filters	X	X	X	X	X
ccbf.so	Internal				X	
cmmmap000.bin	Tables for character mapping (all character sets)	X	X	X	X	X
cmmmap000.sbc	Tables for character mapping (single-byte character sets). This file is located in the /common directory.	X	X	X	X	X
cmmmap000.dbc	Identical to cmmmap000.bin, but renamed for clarity (.dbc = double-byte character). This file is located in the common directory.	X	X	X	X	X
exbf.so	Internal				X	
libfreetype.so.6	TrueType font rendering module for the GD output solution. 32-bit Linux and Solaris Sparc only.	X	X	X	X	X
oitnsf.id	Support file for the vsnsf filter.	X	X	X	X	X
pageml.dtd	The Document Type Definition for the PageML schema				X	
pageml.xsd	The Extensible Schema Definition for the PageML schema				X	

Library	Description	HTML Export	Image Export	PDF Export	Search Export	XML Export
searchml3.dtd	The Document Type Definitions for the SearchML schema				X	
searchml3.xsd	The Extensible Schema Definitions for the SearchML schema				X	

3.3 The Basics

Sample applications are provided with the SDK. These applications demonstrate most of the concepts described in this manual. For a complete description of the sample applications, see [Chapter 8, "Sample Applications."](#)

3.3.1 What You Need in Your Source Code

Any source code that uses this product should `#include` the file `sccex.h` and `#define` `UNIX`. For example, a 32-bit UNIX application might have a source file with the following lines:

```
#define UNIX
#include <sccex.h>
```

A 64-bit UNIX application might have a source file with the following lines:

```
#define UNIX
#define UNIX_64
#include <sccex.h>
```

3.3.2 Information Storage

This software is based on the Oracle Outside In Viewer Technology (or simply "Viewer Technology"). A file of default options is always created, and a list of available filters and a list of available display engines are also built by the technology, usually the first time the product runs (for UNIX implementations). You do not need to ship these lists with your application.

Lists are stored in the `$HOME/.oit` directory. If the `$HOME` environment variable is not set, the files are put in the same directory as the Oracle Outside In Technology. If a `/.oit` directory does not exist in the user's `$HOME` directory, the `.oit` directory is created automatically by the technology. The files are automatically regenerated if corrupted or deleted.

The files are:

- n *.f: Filter lists
- n *.d: Display engine list
- n *.opt: Persistent options

The technology does not actually use the list of default options created by the Viewer Technology.

The filenames are intended to be unique enough to avoid conflict for any combination of machine name and install directory. This is intended to prevent problems with version conflicts when multiple versions of the Viewer Technology and/or other Viewer Technology-based products are installed on a single system. The filenames are built from an 11-character string derived from the directory the Oracle Outside In Technology resides in and the name of the

machine it is being run on. The string is generated by code derived from the RSA Data Security, Inc. MD5 Message-Digest Algorithm.

The products still function if these files cannot be created for some reason. In that situation, however, significant performance degradation should be expected.

3.4 Character Sets

The strings passed in the UNIX API are ISO8859-1 by default.

To optimize performance on systems that do not require DBCS support, a second character mapping bin file, that does not contain any of the DBCS pages, is now included. The second bin file gives additional performance benefits for English documents, but cannot handle DBCS documents. To use the new bin file, replace the `cmmmap000.bin` with the new bin file, `cmmmap000.sbc`. For clarity, a copy of the `cmmmap000.bin` file (`cmmmap000.dbc`) is also included. Both `cmmmap000.sbc` and `cmmmap000.dbc` are located in the `/sdk/common` directory of the technology.

Note: All of the Search Export flavors produce most text in UTF-8 encoded Unicode. Two exceptions to this are the characters in `<unmapped>` elements and XMP metadata (which is passed through without character mapping being applied).

3.5 Runtime Considerations

The following is information to consider during run-time.

3.5.1 X Server Requirement

Note: The X Server requirement can be eliminated by setting the `SCCOPT_RENDERING_PREFER_OIT` option to `TRUE`. When using Search Export, this option is valid when the product is set to produce PageML output.

For Search Export, a X-Server is only necessary when using the PageML format. Access to a running X Windows server and the presence of Motif (or LessTif on Linux) are required to convert from vector formats on UNIX systems. Examples of vector graphics files include CAD drawings and presentation files such as Power Point 97 files.

A runtime check for the presence of X libraries is performed to accommodate system with and without available X servers. This check looks on the system-specific library path variable for the X libraries. If the X libraries are not found, this product does not perform vector graphics conversion.

Be sure to set the `$DISPLAY` environment variable before running this product when non-raster/vector graphic conversion is needed. This is especially important to remember in situations such as CGI programs that start with a limited environment.

For example, when running the technology from a remote session, setting `DISPLAY=:0.0` tells the system to use the X Windows server on the console.

3.5.2 OLE2 Objects

Some documents that the developer is attempting to convert may contain embedded OLE2 objects. There are platform-dependent limits on what the technology can do with OLE2 objects. However, Oracle Outside In attempts to take advantage of the fact that some documents accompany an OLE2 embedding with a graphic "snapshot," in the form of a Windows metafile.

On all platforms, when a metafile snapshot is available, the technology uses it to convert the object. When a metafile snapshot is not available on UNIX platforms, the technology is unable to convert the OLE2 object.

3.5.3 Signal Handling

These products trap and handle the following signals:

- n SIGABRT
- n SIGBUS
- n SIGFPE
- n SIGILL
- n SIGINT
- n SIGSEGV
- n SIGTERM

Developers who wish to override our default handling of these signals should set up their own signal handlers. This may be safely done after the developer's application has called `DAInitEx()`.

Note: The Java Native Interface (JNI) allows Java code to call and be called by native code (C/C++ in the case of OIT). You may run into problems if Java isn't allowed to handle signals and forward them to OIT. If OIT catches the signals and forwards them to Java, the JVMs will sometimes crash. OIT installs signal handlers when `DAInitEx()` is called, so if you call OIT after the JVM is created, you will need to use `libjsig`. Refer here for more information:

<http://www.oracle.com/technetwork/java/javase/index-137495.html>

3.5.4 Runtime Search Path and \$ORIGIN

Libraries and sample applications are all built with the `$ORIGIN` variable as part of the binaries' runtime search path. This means that at runtime, OIT libraries will automatically look in the directory they were loaded from to find their dependent libraries. You don't necessarily need to include the technology directory in your `LD_LIBRARY_PATH` or `SHLIB_PATH`.

As an example, an application that resides in the same directory as the OIT libraries and includes `$ORIGIN` in its runtime search path will have its dependent OIT libraries found automatically. You will still need to include the technology directory in your linker's search path at link time using something like `-L` and possibly `-rpath-link`.

Another example is an application that loads OIT libraries from a known directory. The loading of the first OIT library will locate the dependent libraries.

Note: This feature does not work on AIX and FreeBSD.

3.6 Environment Variables

Several environment variables may be used at run time. Following is a short summary of those variables and their usage.

Variable	Description
\$LD_LIBRARY_PATH (FreeBSD, HP-UX Itanium 64, Linux, Solaris) \$SHLIB_PATH (HP-UX RISC 32) \$LIBPATH (AIX, iSeries)	These variables help your system's dynamic loader locate objects at runtime. If you have problems with libraries failing to load, try adding the path to the Oracle Outside In libraries to the appropriate environment variable. See your system's manual for the dynamic loader and its configuration for details. Note that for products that have a 64-bit PA/RISC, 64-bit Solaris and Linux PPC/PPC64 distributable, they will also go under \$LD_LIBRARY_PATH.
\$DISPLAY	Must be set to point to a valid X Server to render files, unless you plan to use the SCCOPT_RENDERING_PREFER_OIT option. For details, see Section 3.5.1, "X Server Requirement."
\$GDFONTPATH	Must be set if you intend to use the SCCOPT_RENDERING_PREFER_OIT option. This variable includes one or more paths to fonts for use with Oracle Outside In's internal graphics rendering code.
\$HOME	Must be set to allow the system to write the option, filter and display engine lists. For details, see Section 3.3.2, "Information Storage."

3.7 Changing Resources

All of the strings used in the UNIX versions of Oracle Outside In products are contained in the lodlstr.h file. This file, located in the resource directory, can be modified for internationalization and other purposes. Everything necessary to rebuild the resource library to use the modified source file is included with the SDK.

In addition to lodlstr.h, the scclo.o object file is provided. This is necessary for the linking phase of the build. A makefile has also been provided for building the library. The makefile allows building on all of the UNIX platforms supported by Oracle Outside In. It may be necessary to make minor modifications to the makefile so the system header files and libraries can be found for compiling and linking.

Standard INCLUDE and LIB *make* variables are defined for each platform in the makefile. Edit these variables to point to the header files and libraries on your particular system. Other make variables are:

- TECHINCLUDE: May need to be edited to point to the location of the Oracle Outside In /common header files supplied with the SDK.
- BUILDDIR: May need to be edited to point to the location of the makefile, lodlstr.h, and scclo.o (which should all be in the same directory).

After these variables are set, change to the build directory and type make. The libsc_lo resource library is built and placed in the appropriate platform-specific directory. To use this library, copy it into the directory where the Oracle Outside In product is stored and the new, modified resource strings are used by the technology.

Menu constants are included in lomenu.h in the common directory.

3.8 HP-UX Compiling and Linking

The `libsc_ex.so` and `libsc_da.so` libraries are the only ones that must be linked with your application. They can be loaded when your application starts by linking them directly at compile time or they can be loaded dynamically by your application using library load functions (for example, `shl_load`).

The shared libraries are dependent on the presence of the X libraries `Xm`, `Xt` and `X11` if vector graphics support is required. It is the application developer's responsibility to ensure that the needed functions from these libraries are present before the product libraries are used.

The following are example command lines used to compile the sample application **exsimple** from the `/sdk/samplecode/unix` directory. The command lines are separated into sections for HP-UX and HP-UX on Itanium (which requires GCC). This command line is only an example. The actual command line required on the developer's system may vary. The example assumes that the include and library file search paths for the technology libraries and any required X libraries are set correctly. If they are not set correctly, the search paths for the include and/or library files must be explicitly specified via the `-I include file path` and/or `-L library file path` options, respectively, so that the compiler and linker can locate all required files.

3.8.1 HP-UX on RISC

```
cc -w -o ../exsimple/unix/exsimple ../exsimple/unix/exsimple.c +DAportable -Ae
-I/usr/include -I../common -L../demo -L/usr/lib -lsc_ex -lsc_da
-Wl,+s,+b,'$ORIGIN'
```

3.8.2 HP-UX on RISC (64 bit)

```
cc -w -o ../exsimple/unix/exsimple ../exsimple/unix/exsimple.c +DD64
-I/usr/include -I../common -L../demo -L/usr/lib/pa20_64 -DUNIX_64 -lsc_ex
-lsc_da -Wl,+s,+b,'$ORIGIN'
```

3.8.3 HP-UX on Itanium (64 bit)

```
cc -w -o ../exsimple/unix/exsimple ../exsimple/unix/exsimple.c +DD64
-I../common -L../demo -lsc_ex -lsc_da -DUNIX_64 -Wl,+s,+b,'$ORIGIN'
```

3.9 IBM AIX Compiling and Linking

All libraries should be installed into a single directory and the directory must be included in the system's shared library path (`$LIBPATH`). `$LIBPATH` *must* be set and must point to the directory containing the Oracle Outside In Technology.

Oracle Outside In Technology has been updated to increase performance, at a cost of using more memory. It is possible that this increased memory usage may cause a problem on AIX systems, which can be very conservative in the amount of memory they grant to processes. If your application experiences problems due to memory limitations with Oracle Outside In, you may be able to fix this problem by using the "large page" memory model.

If you anticipate viewing or converting very large files with Oracle Outside In technology, we recommend linking your applications with the `-bmaxdata` flag. For example:

```
cc -o foo foo.c -bmaxdata:0x80000000
```

If you are currently seeing "illegal instruction" errors followed by immediate program exit, this is likely due to not using the large data model.

The shared libraries are dependent on the presence of the X libraries Xm, Xt and X11 if vector graphics support is required. It is the application developer's responsibility to ensure that the needed functions from these libraries are present before the product libraries are used.

The following is an example command line used to compile the sample application `exsimple` from the `/sdk/samplecode/unix` directory. This command line is only an example. The actual command line required on the developer's system may vary. The example assumes that the include and library file search paths for the technology libraries and any required X libraries are set correctly. If they are not set correctly, the search paths for the include and/or library files must be explicitly specified via the `-I include file path` and/or `-L library file path` options, respectively, so that the compiler and linker can locate all required files. Developers need to pass `-brtl` to the linker to list libraries in the link command as dependencies of their applications.

Developers may need to use the `-qcpluscmt` flag to allow C++ style comments.

3.9.1 IBM AIX (32-bit pSeries)

```
gcc -w -o ../exsimple/unix/exsimple ../exsimple/unix/exsimple.c -I../common
-L../demo -lsc_ex -lsc_da -DFUNCPROTO -Wl, -brtl
```

3.9.2 IBM AIX PPC (64-bit)

```
gcc -w -o ../exsimple/unix/exsimple ../exsimple/unix/exsimple.c -maix64
-I../common -L../demo -lsc_ex -lsc_da -DUNIX_64 -DFUNCPROTO -Wl, -brtl
```

3.10 Linux Compiling and Linking

This section discusses issues involving Linux compiling and linking.

3.10.1 Library Compatibility

This section discusses Linux compatibility issues when using libraries.

3.10.1.1 Motif Libraries

Problems can be seen when using Export products and trying to convert graphics files. For example, zero-byte graphics files are generated if the technology cannot find the proper Motif library. You can check to see if this is the case by running the following command:

```
ldd libos_xwin.so
```

This prints a list of the dependencies that this library has. If the line for the Motif library is similar to the following then your system may not have a compatible Motif library:

```
libXm.so.3 => not found
```

The solution is to install a compatible Motif library and use it to build your application. Often, the installation discs for your particular Linux platform have the proper libraries. If your installation discs do not have the libraries, instructions for downloading a binary rpm can be found at <http://rpmfind.net/linux/RPM>.

If you are doing development, you must use the proper header files, as well.

The following is a list of the Motif library versions used by Oracle when building and testing the Oracle Outside In binaries.

- x86 Linux: OpenMotif v. 2.2.3
- zSeries Linux: OpenMotif v. 2.2.3

- Itanium Linux: OpenMotif v. 2.1.30

If a directory needs to be specified for the compiler to find the shared libraries, the `$LD_LIBRARY_PATH` environment variable is recommended. This prevents the compiler from hard-coding the library's current directory into the executable as the only directory to search for the library at run time. Instead, the system first searches the directories specified by `$LD_LIBRARY_PATH` for the library.

3.10.1.2 GLIBC and Compiler Versions

The following table indicates the compiler version used and the minimum required version of the GNU standard C library needed for Oracle Outside In operation.

Distribution	Compiler Version	GLIBC Version
x86 Linux	3.3.2	libc.so.6 (2.3 or newer)
Itanium Linux	3.3.2	libc.so.6 (2.3 or newer)
zSeries Linux	3.3.6	libc.so.6 (2.3.2 or newer)

3.10.1.3 Other Libraries

In addition to `libc.so.6`, Oracle Outside In is dependent upon the following libraries:

- `libXm.so.3` (in particular, `libXm.so.3.0.2` or newer, due to issues in OpenMotif 2.2.2)
- `libXt.so.6`
- `libstdc++.so.5.0.5`
- `libgcc_s.so.1`

`libgcc_s.so.1` was introduced with GCC 3.0, so any distribution based on a pre-GCC 3.0 compiler does not include `libgcc_s.so.1`.

The following table summarizes what is included with the RedHat and SUSE distributions supported by Oracle Outside In and what needs to be added/modified to make Oracle Outside In run on these systems. Developers may have trouble building with `libstdc++.so.5` versions before 5.0.5 due to unversioned symbols. Upgrade to 5.0.5 to correct the problem.

3.10.1.3.1 Libraries on Linux Systems as Distributed (IA32)

Advanced Server 3.0

Included	To be added
<code>libc.so.6</code> version	<code>/lib/libc-2.3.2</code>
<code>libstdc++</code>	<code>/usr/lib/libstdc++.so.5.0.3</code>
<code>libgcc_s.so.1</code>	<code>/lib/libgcc_s.so-3.2.3-20030829.so.1</code>
<code>libXm.so.X</code>	<code>libXm.so.2</code> (OpenMotif 2.1.30-8) <code>libXm.so.3.0.1</code> (OpenMotif 2.2.2-16)
Required to Use Oracle Outside In	<ul style="list-style-type: none"> • Default system install has the proper <code>libstdc++.so.5</code> • Default system install includes <code>libgcc_s.so.1</code> • Update to <code>>= libXm.so.3.0.2</code> (OpenMotif <code>>=2.2.3</code>) • Install X libraries

Advanced Server 4.0

Included	To be added
libc.so.6 version	/lib/libc-2.3.4
libstdc++	/usr/lib/libstdc++.so.6.0.3
libgcc_s.so.1	/usr/lib/libgcc_s.so-3.4.3-20041213.so.1
libXm.so.X	libXm.so.2 (OpenMotif 2.1.30-11) libXm.so.3.0.2 (OpenMotif 2.2.3-6)
Required to Use Oracle Outside In	<ul style="list-style-type: none"> • Install libstdc++.so.5 (included with gcc 3.2 - 3.3.6) • Default system install includes libgcc_s.so.1 • Install Motif 2.2.3 from distribution media • Install X libraries

SUSE 8.1

Included	To be added
libc.so.6 version	/lib/libc.so.6 (GLIBC 2.2.5)
libstdc++	/usr/lib/libstdc++.so.5.0.0
libgcc_s.so.1	/lib/libgcc_s.so.1
libXm.so.X	libXm.so.3.0.1
Required to Use Oracle Outside In	<ul style="list-style-type: none"> • Default system install has proper libstdc++.so.5 • Default system install has libgcc_s.so.1 • Update to >= libXm.so.3.0.2 (OpenMotif >=2.2.3) • Install X libraries

SUSE 9.0

Included	To be added
libc.so.6 version	/lib/libc.so.6 (GLIBC 2.3.4)
libstdc++	/usr/lib/libstdc++.so.5.0.6 + old libraries
libgcc_s.so.1	/lib/libgcc_s.so.1
libXm.so.X	libXm.so.3.0.1
Required to Use Oracle Outside In	<ul style="list-style-type: none"> • Default system install has proper libstdc++.so.5 • Default system install has libgcc_s.so.1 • Update to >= libXm.so.3.0.2 (OpenMotif >=2.2.3) • Install X libraries

3.10.1.3.2 Libraries on Linux Systems as Distributed (IA64)**SUSE 8.1**

Included	To be added
libc.so.6 version	/lib/libc.so.6 (GLIBC 2.2.5)

Included	To be added
libstdc++	/usr/lib/libstdc++.so.5.0.0
libgcc_s.so.1	/lib/libgcc_s.so.1
libXm.so.X	libXm.so.3.0.1
Required to Use Oracle Outside In	<ul style="list-style-type: none"> • Default system install has proper libstdc++.so.5 • Default system install has libgcc_s.so.1 • Update to >= libXm.so.3.0.2 (OpenMotif >=2.2.3) • Install X libraries

SUSE 9.0

Included	To be added
libc.so.6 version	/lib/libc.so.6 (GLIBC 2.3.4)
libstdc++	/usr/lib/libstdc++.so.5.0.6 + old libraries
libgcc_s.so.1	/lib/libgcc_s.so.1
libXm.so.X	libXm.so.3.0.1
Required to Use Oracle Outside In	<ul style="list-style-type: none"> • Default system install has proper libstdc++.so.5 • Default system install has libgcc_s.so.1 • Update to >= libXm.so.3.0.2 (OpenMotif >=2.2.3) • Install X libraries

SUSE Linux Enterprise Server 8.0

Included	To be added
libc.so.6 version	/lib/libc.so.6.1 (GLIBC 2.2.6)
libstdc++	/usr/lib/libstdc++-libc6.2-2.so.3 /usr/lib/libstdc++.so.5.0.0
libgcc_s.so.1	/lib/libgcc_s.so.1
libXm.so.X	libXm.so.3.0.1
Required to Use Oracle Outside In	<ul style="list-style-type: none"> • Default system install has proper libstdc++.so.5. • Default system install has libgcc_s.so.1 • Update to >= libXm.so.3.0.2 (OpenMotif >=2.2.3) • Install X libraries

3.10.2 Compiling and Linking

The `libsc_ex.so` and `libsc_da.so` are the only libraries that must be linked with your applications. They can be loaded when your application starts by linking them directly at compile time or they can be loaded dynamically by your application using library load functions (for example, `dlopen`).

The shared libraries are dependent on the presence of the X libraries `Xm`, `Xt` and `X11` if vector graphics support is required. It is the application developer's responsibility to ensure that the needed functions from these libraries are present before the product libraries are used.

The following are example command lines used to compile the sample application `exsimple` from the `/sdk/samplecode/unix` directory. This command line is only an example. The actual command line required on the developer's system may vary.

The example assumes that the include and library file search paths for the technology libraries and any required X libraries are set correctly. If they are not set correctly, the search paths for the include and/or library files must be explicitly specified via the `-I include file path` and/or `-L library file path` options, respectively, so the compiler and linker can locate all required files.

The `-L/usr/X11R6/lib` option is also available.

3.10.2.1 Linux 32-bit, including Linux PPC

```
gcc -w -o ../exsimple/unix/exsimple ../exsimple/unix/exsimple.c
-I/usr/local/include -I../common -L../demo -L/usr/local/lib -lsc_ex -lsc_da
-Wl,-rpath,../demo -Wl,-rpath,'${ORIGIN}'
```

3.10.2.2 Linux 64-bit

```
gcc -w -o ../exsimple/unix/exsimple ../exsimple/unix/exsimple.c
-I/usr/local/include -I../common -L../demo -L/usr/local/lib -lsc_ex -lsc_da
-DUNIX_64 -Wl,-rpath,../demo -Wl,-rpath,'${ORIGIN}'
```

3.10.2.3 Linux zSeries

```
gcc -w -o ../exsimple/unix/exsimple ../exsimple/unix/exsimple.c
-I/usr/local/include -I../common -L../demo -L/usr/local/lib -lsc_ex -lsc_da
-Wl,-rpath,../demo -Wl,-rpath,'${ORIGIN}'
```

3.11 Oracle Solaris Compiling and Linking

Note: These products do not support the "Solaris BSD" mode.

All libraries should be installed into a single directory. The `libsc_ex.so`, and `libsc_da.so` libraries must be linked with your application. It can be loaded when your application starts by linking them directly at compile time or they can be loaded dynamically by your application using library load functions (for example, `dlopen`).

The shared libraries are dependent on the presence of the X libraries `Xm`, `Xt` and `X11` if vector graphics support is required. It is the application developer's responsibility to ensure that the needed functions from these libraries are present before the product libraries are used.

The following is an example command line used to compile the sample application `exsimple` from the `/sdk/samplecode/unix` directory. This command line is only an example. The actual command line required on the developer's system may vary. The example assumes that the include and library file search paths for the technology libraries and any required X libraries are set correctly. If they are not set correctly, the search paths for the include and/or library files must be explicitly specified via the `-I include file path` and/or `-L library file path` options, respectively, so that the compiler and linker can locate all required files.

Developers may need to use the `-xccc` flag to allow C++ style comments.

3.11.1 Oracle Solaris SPARC

```
cc -w -o ../exsimple/unix/exsimple ../exsimple/unix/exsimple.c -I/usr/include
-I/usr/dt/share/include -I../common -L../demo -L/usr/lib -L/lib -lsc_ex
-lsc_da -Wl,-R,../demo -Wl,-R,'${ORIGIN}'
```

When running the 32-bit SPARC binaries on Solaris 9 systems, you may see the following error:

```
ld.so.1: simple: fatal: libm.so.1: version `SUNW_1.1.1' not found
(required by file ./libsc_vw.so)
```

This is due to a missing system patch. Please apply one of the following patches (or its successor) to your system to correct.

n For Solaris 9: Patch 111722-04

3.11.2 Oracle Solaris (SPARC) 64

```
cc -w -o ../exsimple/unix/exsimple ../exsimple/unix/exsimple.c -xtarget=generic64
-I/usr/include -I/usr/dt/share/include -I../common -L../demo -L/usr/lib
-L/lib -lsc_ex -lsc_da -DUNIX_64 -Wl,-R,../demo -Wl,-R,'${ORIGIN}'
```

3.11.3 Oracle Solaris x86

Note: Your system will require Solaris patch 108436, which contains the C++ library libCstd.so.1.

```
cc -w -o ../exsimple/unix/exsimple ../exsimple/unix/exsimple.c -I/usr/include
-I/usr/dt/share/include -I../common -L../demo -L/usr/lib -L/lib -lsc_ex
-lsc_da -Wl,-R,../demo -Wl,-R,'${ORIGIN}'
```

3.12 z/OS Compiling and Linking

The `libsc_ex.x` and `libsc_da.x` libraries must be linked with your application. They can be loaded when your application starts by linking them directly at compile time or they can be loaded dynamically by your application using library load functions (for example, `dlopen`).

The shared libraries are dependent on the presence of the X libraries `Xm`, `Xt` and `X11` if vector graphics support is required. It is the application developer's responsibility to ensure that the needed functions from these libraries are present before the product libraries are used.

All libraries should be installed into a single directory and the directory must be included in the system's shared library path (`$LIBPATH`). `$LIBPATH` *must* be set and must point to the directory containing the Oracle Outside In technology.

The following is an example command line used to compile the sample application **exsimple** from the `/sdk/samplecode/unix` directory. This command line is only an example. The actual command line required on the developer's system may vary. The example assumes that the include and library file search paths for the technology libraries and any required X libraries are set correctly. If they are not set correctly, the search paths for the include and/or library files must be explicitly specified via the `-I include file path` and/or `-L library file path` options, respectively, so the compiler and linker can locate all required files.

```
c89 -o ../exsimple/unix/exsimple -I/usr/include/X11 -I/usr/local/include
-I../common -W 'c,ASCII,LANGlvl(ANSI,LANGLONG)' -D_ZOS_SOURCE -D_XOPEN_
SOURCE=500 -Wl,DLL,XPLINK -L../demo -L/usr/local/lib -L/usr/local/lib/oivt
../demo/libsc_fa.x ../demo/libsc_ex.x ../demo/libsc_da.x
../exsimple/unix/exsimple.c
```

Data Access Common Functions

The Data Access module is common to all Oracle Outside In technologies. It provides a way to open a generic handle to a source file. This handle can then be used in the functions described in this chapter.

This chapter includes the following sections:

- n [Section 4.1, "Deprecated Functions"](#)
- n [Section 4.2, "DAInitEx"](#)
- n [Section 4.3, "DADeInit"](#)
- n [Section 4.4, "DAOpenDocument"](#)
- n [Section 4.5, "DACloseDocument"](#)
- n [Section 4.6, "DARetrieveDocHandle"](#)
- n [Section 4.7, "DASetOption"](#)
- n [Section 4.8, "DAGetOption"](#)
- n [Section 4.9, "DAGetFileId"](#)
- n [Section 4.10, "DAGetFileIdEx"](#)
- n [Section 4.11, "DAGetErrorString"](#)
- n [Section 4.12, "DAGetObjectInfo"](#)
- n [Section 4.13, "DAGetTreeCount"](#)
- n [Section 4.14, "DAGetTreeRecord"](#)
- n [Section 4.15, "DAOpenTreeRecord"](#)
- n [Section 4.16, "DAOpenRandomTreeRecord"](#)
- n [Section 4.17, "DASaveInputObject"](#)
- n [Section 4.18, "DASaveTreeRecord"](#)
- n [Section 4.19, "DASaveRandomTreeRecord"](#)
- n [Section 4.20, "DACloseTreeRecord"](#)
- n [Section 4.21, "DASetStatCallback"](#)
- n [Section 4.22, "DASetFileAccessCallback"](#)

4.1 Deprecated Functions

DAInit and DaThreadInit have both been deprecated. DAINitEx now replaces these two functions. All new implementations should use DAINitEX, although the other two functions will continue to be supported.

4.2 DAINitEx

This function tells the Data Access module to perform any necessary initialization it needs to prepare for document access. This function must be called before the first time the application uses the module to retrieve data from any document. This function supersedes the old DAINit and DATHreadInit functions.

Note: DAINitEx should only be called once per application, at application startup time. Any number of documents can be opened for access between calls to DAINitEx and DADeInit. If DAINitEx succeeds, DADeInit must be called regardless of any other API calls.

If the ThreadOption parameter is set to something other than DATHREAD_INIT_NOTHREADS, then this function's preparation includes setting up mutex function pointers to prevent threads from clashing in critical sections of the technology's code. The developer must actually code the threads after this function has been called. DAINitEx should be called only once per process and should be called before the developer's application begins the thread.

Note: Multiple threads are supported for all Windows platforms and the 32-bit versions of Linux x86 and Solaris SPARC. Failed initialization of the threading function will not impair other API calls. If threading isn't initialized or fails, stub functions are called instead of mutex functions.

Prototype

```
DAERR DAINitEx(VTSHORT ThreadOption, VTDWORD dwFlags);
```

Parameters

- n ThreadOption: can be one of the following values:
 - DATHREAD_INIT_NOTHREADS: No thread support requested.
 - DATHREAD_INIT_PTHREADS: Support for PTHREADS requested.
 - DATHREAD_INIT_NATIVETHREADS: Support for native threading requested. Supported only on Microsoft Windows platforms and Oracle Solaris.
- n dwFlags: can be one or more of the following flags OR-ed together
 - OI_INIT_DEFAULT: Options Load and Save are performed normally
 - OI_INIT_NOSAVEOPTIONS: The options file will not be saved on exit
 - OI_INIT_NOLOADOPTIONS: The options file will not be read during initialization.

Return Values

- n DAERR_OK: If the initialization was successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.3 DADeInit

This function tells the Data Access module that it will not be asked to read additional documents, so it should perform any cleanup tasks that may be necessary. This function should be called at application shutdown time, and only if the module was successfully initialized with a call to DAInitEx.

Prototype

```
DAERR DADeInit();
```

Return Values

- DAERR_OK: If the de-initialization was successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.4 DAOpenDocument

Opens a source file to make it accessible by one or more of the data access technologies. If DAOpenDocument succeeds, DACloseDocument must be called regardless of any other API calls.

For IO types other than IOTYPE_REDIRECT, the subdocument specification may be specified as part of the file's path. This is accomplished by appending a question mark delimiter to the path, followed by the subdocument specification. For example, to specify the third item within the file c:\docs\file.zip, specify the path c:\docs\file.zip?item.3 in the call to DAOpenDocument. DAOpenDocument always attempts to open the specification as a file first. In the unlikely event there is a file with the same name (including the question mark) as a file plus the subdocument specification, that file is opened instead of the archive item.

To take advantage of this feature when providing access to the input file using redirected IO, a subdocument specification must be provided via a response to an IOGetInfo message, IOGETINFO_SUBDOC_SPEC. To specify an item in an archive, first follow the standard redirected IO methods to provide a BASEIO pointer to the archive file itself. To specify an item within the archive, a redirected IO object must respond to the IOGETINFO_SUBDOC_SPEC message by copying to the supplied buffer the subdocument specification of the archive item to be opened. This message is received during the processing of DAOpenDocument.

Prototype

```
DAERR DAOpenDocument(
    VTLPDOC    lphDoc,
    VTDWORD    dwSpecType,
    VTLPVOID    pSpec,
    VTDWORD    dwFlags);
```

Parameters

- lphDoc: Pointer to a handle that will be filled with a value uniquely identifying the document to data access. The developer uses this handle in subsequent calls to data access to identify this particular source file. This is not an operating system file handle.
- dwSpecType: Describes the contents of pSpec. Together, dwSpecType and pSpec describe the location of the source file. Must be one of the following values:
 - IOTYPE_ANSIPATH: Windows only. pSpec points to a NULL-terminated full path name using the ANSI character set and FAT 8.3 (Win16) or NTFS (Win32 and Win64) file name conventions.

- IOTYPE_UNICODEPATH: Windows only. pSpec points to a NULL-terminated full path name using the Unicode character set and NTFS (Win32 and Win64) file name conventions.
 - IOTYPE_UNIXPATH: UNIX platforms only. pSpec points to a NULL-terminated full path name using the system default character set and UNIX path conventions. Unicode paths can be accessed on UNIX platforms by using a UTF-8 encoded path with IOTYPE_UNIXPATH.
 - IOTYPE_REDIRECT: All platforms. pSpec points to a developer-defined struct that allows the developer to redirect the IO routines used to read the file. For more information, see [Chapter 6, "Redirected IO."](#)
 - IOTYPE_ARCHIVEOBJECT: All platforms. Opens an embedded archive object for data access. pSpec points to a structure IOSPECARCHIVEOBJECT (see [Section 4.4.2, "IOSPECARCHIVEOBJECT Structure"](#)) that has been filled with values returned in a SCCCA_OBJECT content entry from Content Access.
 - IOTYPE_LINKEDOBJECT: All platforms. Opens an object specified by a linked object for data access. pSpec points to a structure IOSPECLINKEDOBJECT (see [Section 4.4.1, "IOSPECLINKEDOBJECT Structure"](#)) that has been filled with values returned in an SCCCA_BEGINTAG or SCCCA_ENDTAG with a subtype of SCCCA_LINKEDOBJECT content entry from Content Access.
 - IOTYPE_OBJECT: All platforms. Opens an object (archive, embedded, or linked) for data access. pSpec points to a structure SCCDAOBJECT (see [Section 4.4.3, "SCCDAOBJECT Structure"](#)) that has been filled with values from Content Access (SCCCA_OBJECT or SCCCA_BEGINTAG with a subtype of SCCCA_LINKEDOBJECT) or from the <document> element in the SearchML flavor of Search Export.
- n pSpec: File location specification.
- n dwFlags: The low WORD is the file ID for the document (0 by default). If you set the file ID incorrectly, the technology fails. If set to 0, the file identification technology determines the input file type automatically. The high WORD should be set to 0. It may also be set to the following flags:
- DAOPENDOCUMENT_ARCHIVEONLYMODE: This flag may only be used with archive files. It opens the archive in a special mode that is only usable with [DA SaveRandomTreeRecord](#) and [DA OpenRandomTreeRecord](#).
 - DAOPENDOCUMENT_CONTINUEONFAILURE: Some embeddings may have both an OLE representation and an alternate graphic. When this flag is set for IOTYPE_OBJECT, the technology will first try to access the OLE representation. If there are errors, it will then attempt to access the alternate graphic.

Return Values

- n DAERR_OK: Returned if the open was successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.4.1 IOSPECLINKEDOBJECT Structure

Structure used by DAOpenDocument.

Prototype

```
typedef struct IOSPECLINKEDOBJECTtag
{
    VTDWORD    dwStructSize;
```



```

VTSYSPARAM hDoc;
VTDWORD   dwObjectId; /* Object identifier. */
VTDWORD   dwType;     /* Linked Object type */
                /* (SO_LOCATOR_TYPE_*) */
VTDWORD   dwParam1;  /* parameter for DoSpecial call */
VTDWORD   dwParam2;  /* parameter for DoSpecial call */
VTDWORD   dwReserved1; /* Reserved. */
VTDWORD   dwReserved2; /* Reserved. */
} IOSPECLINKEDOBJECT, * PIOSPECLINKEDOBJECT;

```

4.4.2 IOSPECARCHIVEOBJECT Structure

Structure used by DAOpenDocument.

Prototype

```

typedef struct IOSPECARCHIVEOBJECTtag
{
    VTDWORD dwStructSize;
    VTDWORD hDoc; /* Parent Doc hDoc */
    VTDWORD dwNodeId; /* Node ID */
    VTDWORD dwStreamId;
    VTDWORD dwReserved1; /* Must always be 0 */
    VTDWORD dwReserved2; /* Must always be 0 */
} IOSPECARCHIVEOBJECT, * PIOSPECARCHIVEOBJECT;

```

4.4.3 SCCDAOBJECT Structure

Structure used by DAOpenDocument.

Prototype

```

typedef struct SCCDAOBJECTtag
{
    VTDWORD dwSize; /* sizeof(SCCDAOBJECT) */
    VTHDOC hDoc; /* DA handle for the document
                containing the object */
    VTDWORD dwObjectType; /* SCCCA_EMBEDDED_OBJECT,
                SCCCA_LINKED_OBJECT,
                SCCCA_COMPRESSED_FILE or
                SCCCA_ATTACHMENT */
    VTDWORD dwData1; /* Data identifying the object */
    VTDWORD dwData2; /* Data identifying the object */
    VTDWORD dwData3; /* Data identifying the object */
    VTDWORD dwData4; /* Data identifying the object */
} SCCDAOBJECT, * PSCCDAOBJECT;

```

4.5 DACloseDocument

This function is called to close a file opened by the reader that has not encountered a fatal error.

Prototype

```

DAERR DACloseDocument(
    VTHDOC hDoc);

```

Parameters

- n hDoc: Identifier of open document. Must be a handle returned by the DAOpenDocument function.

Return Value

- DAERR_OK: Returned if close succeeded. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.6 DARetrieveDocHandle

This function returns the document handle associated with any type of Data Access handle. This allows the developer to only keep the value of hItem, instead of both hItem and hDoc.

Prototype

```
DAERR DARetrieveDocHandle(  
    VTHDOC    hItem,  
    VTLPHDOC  phDoc);
```

Parameters

- hItem: Identifier of open document. May be the subhandle returned by the DAOpenDocument or DAOpenTreeRecord functions in the data access submodule. Passing in an hDoc created by DAOpenDocument for this parameter results in an error.
- phDoc: Pointer to a handle to be filled with the document handle associated with the passed subhandle.

Return Value

- DAERR_OK: Returned if the handle in phDoc is valid. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.7 DASetOption

This function is called to set the value of a data access option.

Prototype

```
DAERR DASetOption(  
    VTHDOC    hDoc,  
    VTDWORD   dwOptionId,  
    VTLPVOID  pValue,  
    VTDWORD   dwValueSize);
```

Parameters

- hDoc: Identifier of open document. May be a VTHDOC returned by the DAOpenDocument function, or the subhandle returned by the DAOpenDocument or DAOpenTreeRecord functions (VTHCONTENT, VTHTEXT, etc.). Setting an option for a VTHDOC affects all subhandles opened under it, while setting an option for a subhandle affects only that handle.

If this parameter is NULL, then setting the option affects all documents opened thereafter. Once an option is set using the NULL handle, this option becomes the default option thereafter. This parameter should only be set to NULL if the option being set can take that value.

- dwOptionId: The identifier of the option to be set.
- pValue: Pointer to a buffer containing the value of the option.
- dwValueSize: The size in bytes of the data pointed to by pValue. For a string value, the NULL terminator should be included when calculating dwValueSize.

Return Value

- DAERR_OK: Returned if DASEToption succeeded. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.8 DAGetOption

This function is called to retrieve the value of a data access option. The results of a call to this option are only valid if DASEToption has already been called on the option.

Prototype

```
DAERR DAGetOption(
    VTHDOC    hItem,
    VTDWORD   dwOptionId,
    VTLPVOID  pValue,
    VTLPDWORD pSize);
```

Parameters

- hItem: Identifier of open document. May be a VTHDOC returned by the DAOpenDocument function, or the subhandle returned by the DAOpenDocument or DAOpenTreeRecord functions (VTHCONTENT, VTHTEXT, etc.). Getting an option for a VTHDOC gets the value of that option for that handle, which may be different than the subhandle's value.
- dwOptionId: The identifier of the option to be returned.
- pValue: Pointer to a buffer containing the value of the option.
- pSize: This VTDWORD should be initialized by the caller to the size of the buffer pointed to by pValue. If this size is sufficient, the option value is copied into pValue and pSize is set to the actual size of the option value. If the size is not sufficient, pSize is set to the size of the buffer needed for the option and an error is returned.

Return Value

- DAERR_OK: Returned if DAGetOption was successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.9 DAGetFileId

This function allows the developer to retrieve the format of the file based on the technology's content-based file identification process. This can be used to make intelligent decisions about how to process the file and to give the user feedback about the format of the file they are working with.

Note: in cases where File ID returns a value of FI_UNKNOWN, then this function will apply the Fallback Format before returning a result.

Prototype

```
DAERR DAGetFileId(
    VTHDOC    hDoc,
    VTLPDWORD pdwFileId);
```

Parameters

- hDoc: Identifier of open document. May be a VTHDOC returned by the DAOpenDocument function, or the subhandle returned by the DAOpenDocument or DAOpenTreeRecord functions (VTHEXPORT, VTHCONTENT, VTHTEXT, etc.).

- n pdwFileId: Pointer to a DWORD that receives a file identification number for the file. These numbers are defined in sccfi.h.

Return Value

- n DAERR_OK: Returned if DAGetFileId was successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.10 DAGetFileIdEx

This function allows the developer to retrieve the format of the file based on the technology's content-based file identification process. This can be used to make intelligent decisions about how to process the file and to give the user feedback about the format of the file they are working with. This function has all the functionality of DAGetFileID and adds the ability to return the raw FI value; in other words, the value returned by normal FI, without applying the FallbackFI setting.

Prototype

```
DAERR DAGetFileIdEx(
    VTHDOC      hDoc,
    VTLPDWORD   pdwFileId,
    VTDWORD     dwFlags);
```

Parameters

- n hDoc: Identifier of open document. May be a VTHDOC returned by the DAOpenDocument function, or the subhandle returned by the DAOpenDocument or DAOpenTreeRecord functions (VTHEXPORT, VTHCONTENT, VTHTEXT, etc.).
- n pdwFileId: Pointer to a DWORD that receives a file identification number for the file. These numbers are defined in sccfi.h.
- n dwFlags: DWORD that allows user to request specific behavior.
 - DA_FILEINFO_RAWFI: This flag tells DAGetFileIdEx() to return the result of the File Identification operation before Extended File Ident. is performed and without applying the FallbackFI value.

Return Value

- n DAERR_OK: Returned if DAGetFileIdEx was successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned. See the following tables for examples of expected output depending on the value of various options.

Values with RAWFI turned off

Input file type	ExtendedFI	FallbackID	DAGetFileId	DAGetFileIdEx
true binary	off	fallback value	fallback value	fallback value
true binary	on	fallback value	fallback value	fallback value
true text	off	fallback value	fallback value	fallback value
true text	on	fallback value	40XX	40XX

Values with RAWFI turned on

Input file type	ExtendedFI	FallbackID	DAGetFileId	DAGetFileIdEx
true binary	off	fallback value	fallback value	1999
true binary	on	fallback value	fallback value	1999
true text	off	fallback value	fallback value	1999
true text	on	fallback value	40XX	1999

4.11 DAGetErrorString

This function returns to the developer a string describing the input error code. If the error string returned does not fit the buffer provided, it is truncated.

```

VTVOID DAGetErrorString(
    DAERR      deError,
    VTLPVOID   pBuffer,
    VTDWORD    dwBufSize);

```

Parameters

- **deError**: Error code passed in by the developer for which an error message is to be returned.
- **pBuffer**: This buffer is allocated by the caller and is filled in with the error message by this routine. The error message will be a NULL-terminated string.
- **dwBufSize**: Size of what pBuffer points to in bytes.

Return Value

- none

4.12 DAGetObjectInfo

This function returns information about the document or object pointed to by hDoc. The object may be an embedded object, a linked object, or a compressed file.

```

DAERR DAGetObjectInfo(
    VTHDOC     hDoc,
    VTDWORD    dwInfoId,
    VTLPVOID   pInfo);

```

Parameters

- **hDoc**: The handle returned by DAOpenDocument.
- **dwInfoId**: The identifier of the requested information. Can be any of the following values:
 - **DAOBJECT_NAME_A**: Retrieves the name of the object, in 8-bit characters. pInfo points to a buffer of size DA_PATHSIZE.
 - **DAOBJECT_NAME_W**: Retrieves the name of the object in Unicode characters. pInfo points to a buffer of 16 bit characters of size DA_PATHSIZE.
 - **DAOBJECT_FORMATID**: Retrieves the file ID of the object. pInfo points to a VTDWORD value.
 - **DAOBJECT_COMPRESSIONTYPE**: Retrieves an identifier of the type of compression used to store the object, if known. pInfo points to a VTDWORD value.

- **DAOBJECT_FLAGS**: Retrieves a bitfield of flags indicating additional attributes of the object. pInfo points to a VTDWORD value. Possible flag values include **DAOBJECTFLAG_PARTIALFILE** (would not normally exist outside the source document), **DAOBJECTFLAG_PROTECTEDFILE** (encrypted or password protected), **DAOBJECTFLAG_LINKTOFILE** (indicates that an OLE object is linked to the file and a corresponding file is not found on the host machine), **DAOBJECTFLAG_UNIDENTIFIEDFILE** (indicates that an object could not be identified), and **DAOBJECTFLAG_UNSUPPORTEDCOMP** (compressed with an unsupported compression), and **DAOBJECTFLAG_ARCKNOWNCRYPT** (see note below).
- n **pInfo**: Destination of the requested information. The possible types are described in the preceding section about dwInfoId.

Note: **DAOBJECTFLAG_ARCKNOWNCRYPT** indicates that the object is protected by a known encryption. It can be accessed after the correct credentials (password and/or Lotus Notes id file) are provided through the File Access Callback. For details, see [Section 4.22](#), "[DASetFileAccessCallback](#)."

Return Values

- n **DAERR_OK**: Returned if the save was successful. Otherwise, one of the other **DAERR_** values in `sccda.h` or one of the **SCCERR_** values in `scerr.h` is returned.

4.13 DAGetTreeCount

This function is called to retrieve the number of records in an archive file.

```
DAERR DAGetTreeCount(  
    VTHDOC      hDoc,  
    VTLPDWORD   lpRecordCount);
```

Parameters

- n **hDoc**: Identifier of open document. May be a **VTHDOC** returned by the **DAOpenDocument** function, or the subhandle returned by any of the **DAOpenDocument** or **DAOpenTreeRecord** functions (**VTHCONTENT**, **VTHTEXT**, etc.).
- n **lpRecordCount**: A pointer to a **VTLPDWORD** that is filled with the number of stored archive records.

Return Value

- n **DAERR_OK**: **DAGetTreeCount** was successful. Otherwise, one of the other **DAERR_** values in `sccda.h` or one of the **SCCERR_** values in `scerr.h` is returned.
- n **DAERR_BADPARAM**: The selected file does not contain an archive section, or the requested record does not exist.

4.14 DAGetTreeRecord

This function is called to retrieve information about a record in an archive file.

```
DAERR DAGetTreeRecord(  
    VTHDOC      hDoc,  
    PSCCDATREENODE pTreeNode);
```

Parameters

- hDoc: Identifier of open document. May be a VTHDOC returned by the DAOpenDocument function, or the subhandle by any of the DAOpenDocument or DAOpenTreeRecord functions (VTHCONTENT, VTHTEXT, etc.).
- pTreeNode: A pointer to a PSCCDATREENODE structure that is filled with information about the selected record.

Return Values

- DAERR_OK: DAGetTreeRecord was successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.
- DAERR_BADPARAM: The selected file does not contain an archive section, or the requested record does not exist.
- DAERR_EMPTYFILE: Empty file.
- DAERR_PROTECTEDFILE: Password protected or encrypted file.
- DAERR_SUPFILEOPENFAILS: Supplementary file open failed.
- DAERR_FILTERNOTAVAIL: The file's type is known, but the appropriate filter is not available.
- DAERR_FILTERLOADFAILED: An error occurred during the initialization of the appropriate filter.

4.14.1 SCCDATREENODE Structure

This structure is passed by the OEM through the DAGetTreeRecord function. The structure is defined in sccda as follows:

```
typedef struct SCCDATREENODEtag{
    VTDWORD    dwSize;
    VTDWORD    dwNode;
    VTBYTE     szName[1024];
    VTDWORD    dwFileSize;
    VTDWORD    dwTime;
    VTDWORD    dwFlags;
    VTDWORD    dwCharSet;
} SCCDATREENODE, *PSCCDATREENODE;
```

Parameters

- dwSize: Must be set by the OEM to sizeof(SCCDATREENODE).
- dwNode: The number of the record to retrieve information about. The first node is node 0.
- szName: A buffer to hold the name of the record.
- dwFileSize: Returns the file size, in bytes, of the requested record.
- dwTime: Returns the timestamp of the requested record, in MS-DOS time.
- dwFlags: Returns additional information about the node. It can be a combination of the following:
 - SCCDA_TREENODEFLAG_FOLDER: Indicating that the selected node is a folder and not a file.
 - SCCDA_TREENODEFLAG_SELECTED: Indicating that the node is selected.
 - SCCDA_TREENODEFLAG_FOCUS: Indicating that the node has focus.

- SCCDA_TREENODEFLAG_ENCRYPT: Indicating that the node is encrypted and can not be decrypted.
 - SCCDA_TREENODEFLAG_ARCKNOWNCRYPT: indicating that the node is encrypted with an unknown encryption and can not be decrypted.
 - SCCDA_TREENODEFLAG_BUFFEROVERFLOW: the name of the node was too long for the szName field.
- dwCharSet: Returns the SO_* (charsets.h) character set of the characters in szName. The output character set is either the default native environment character set or Unicode if the SCCOPT_SYSTEMFLAGS option is set to SCCVW_SYSTEM_UNICODE.

4.15 DAOpenTreeRecord

This function is called to open a record within an archive file and make it accessible by one or more of the data access technologies.

Search Export Only: Search Export's default behavior is to automatically open and process the contents of an archive. Use DAOpenTreeRecord and SCCOPT_XML_SEARCHML_FLAGS to change the default behavior if discrete processing of each document in an archive is desired.

```
DAERR DAOpenTreeRecord(  
    VTHDOC      hDoc,  
    VTLPDOC     lphDoc,  
    VTDWORD     dwRecord);
```

lphDoc is *not* a file handle.

Parameters

- hDoc: Identifier of open document. May be a VTHDOC returned by the DAOpenDocument function, or the subhandle returned by the DAOpenDocument or DAOpenTreeRecord functions (VTHCONTENT, VTHTEXT, etc.).
- lphDoc: Pointer to a handle that is filled with a value uniquely identifying the document to data access. The developer uses this handle in subsequent calls to data access to identify this particular document.
- dwRecord: The record in the archive file to be opened.

Return Value

- DAERR_OK: Returned if DAOpenTreeRecord was successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.16 DAOpenRandomTreeRecord

This function is called to open a record within an archive file and make it accessible by one or more of the data access technologies. It is similar to DAOpenTreeRecord, except that instead of reading the data for all nodes in the archive in a sequential order, this function will only read the data for the requested nodes from the archive. To use this function, you must first process the archive with Content Access or Search Export and save the Node Locator data for later use in this function.

```
DAERR DAOpenRandomTreeRecord(  
    VTHDOC      hDoc,  
    VTLPDOC     lphDoc,  
    SOTREENODELOCATOR sTreeNodeLocator )
```

lphDoc is not a file handle.

Parameters

- **hDoc**: Identifier of open document. This hDoc must come from an archive document opened with `DAOpenDocument` with the flag `DAOPENDOCUMENT_ARCHIVEONLYMODE` set.
- **lphDoc**: Pointer to a handle that will be filled with a value that uniquely identifies the document to data access. The developer will use this handle in subsequent calls to data access to identify this particular document.
- **sTreeNodeLocator**: An `SOTREENODELOCATOR` structure which contains data identifying the desired node. This data should come from a previous conversion of the archive document using Content Access or Search Export.

Return Value

- **DAERR_OK**: Returned if `DAOpenRandomTreeRecord` was successful. Otherwise, one of the other `DAERR_` values in `scda.h` or one of the `SCCERR_` values in `scerr.h` is returned.

4.16.1 DATREENODELOCATOR

```
typedef struct DATREENODELOCATORtag
{
    VTDWORD dwSize; /* size of this structure */
    VTDWORD dwSpecialFlag; /* special flags coming from CA or SX */
    VTDWORD dwData1; /* dwData1 coming from CA or SX */
    VTDWORD dwData2; /* dwData2 coming from CA or SX */
}SCCDATREENODELOCATOR, *PSCCDATREENODELOCATOR;
```

4.16.2 SCCCA_TREENODELOCATOR: Tree Node Locator

This content type contains information to be used in the `SOTREENODELOCATOR` structure, which is used by [DAOpenRandomTreeRecord](#) and [DASaveRandomTreeRecord](#).

4.17 DASaveInputObject

This function saves a copy of the document or object pointed to by `hDoc`. The object may be an embedded object, a linked object or a compressed file.

Some file formats store only partial files as embedded objects. Oracle Outside In is not able to create readable files from these objects. You should use `DAGetObjectInfo` with `dwInfoFold` set to `DAOBJECT_FLAGS` to discern which objects Oracle Outside In can successfully extract.

```
DAERR DASaveInputObject(
    VTHDOC    hDoc,
    VTDWORD   dwSpecType,
    VTLPVOID  pSpec,
    VTDWORD   dwFlags);
```

Parameters

- **hDoc**: The handle returned by `DAOpenDocument`.
- **dwSpecType**: Describes the contents of `pSpec`. Together, `dwSpecType` and `pSpec` describe the location of the source file to which the file will be extracted. Must be one of the following values:
 - **IOTYPE_ANSIPATH**: Windows only. `pSpec` points to a NULL-terminated full path name using the ANSI character set and FAT 8.3 (Win16) or NTFS (Win32 and Win64) filename conventions.

- IOTYPE_REDIRECT: Specifies that redirected I/O will be used to save the file.
 - IOTYPE_UNICODEPATH: Windows only. pSpec points to a NULL-terminated full path name using the Unicode character set and NTFS (Win32 and Win64) file name conventions.
 - IOTYPE_UNIXPATH: X Windows on UNIX platforms only. pSpec points to a NULL-terminated full path name using the system default character set and UNIX path conventions. Unicode paths can be accessed on UNIX platforms by using a UTF-8 encoded path with IOTYPE_UNIXPATH.
- pSpec: File location specification.
 - dwFlags: Currently not used. Should be set to 0.

Return Values

- DAERR_OK: Returned if the save was successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.18 DASaveTreeRecord

This function is called to extract a record in an archive file to disk.

```
DAERR DASaveTreeRecord(  
    VTHDOC      hDoc,  
    VTDWORD     dwRecord,  
    VTDWORD     dwSpecType,  
    VTLPVOID    pSpec,  
    VTDWORD     dwFlags);
```

Parameters

- hDoc: Handle that uniquely identifies the document to data access. This is not an operating system file handle.
- dwRecord: The record in the archive file to be extracted.
- dwSpecType: Describes the contents of pSpec. Together, dwSpecType and pSpec describe the location of the source file to which the file will be extracted. Must be one of the following values:
 - IOTYPE_ANSIPATH: Windows only. pSpec points to a NULL-terminated full path name using the ANSI character set and FAT 8.3 (Win16) or NTFS (Win32 and Win64) filename conventions.
 - IOTYPE_REDIRECT: Specifies that redirected I/O will be used to save the file.
 - IOTYPE_UNICODEPATH: Windows only. pSpec points to a NULL-terminated full path name using the Unicode character set and NTFS (Win32 and Win64) file name conventions.
 - IOTYPE_UNIXPATH: X Windows on UNIX platforms only. pSpec points to a NULL-terminated full path name using the system default character set and UNIX path conventions. Unicode paths can be accessed on UNIX platforms by using a UTF-8 encoded path with IOTYPE_UNIXPATH.
- pSpec: File location specification. See the descriptions for individual dwSpecType values.
- dwFlags: Currently not used. Should be set to 0.

Return Values

- DAERR_OK: Returned if the save was successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.
 - DAERR_UNSUPPORTEDCOMP: Unsupported Compression Encountered.
 - DAERR_PROTECTEDFILE: The file is encrypted.
 - DAERR_BADPARAM: The request option is invalid. The record is possibly a directory.
- Otherwise, one of the other DAERR_ values in sccda.h is returned.

Note: Currently, only extracting a single file is supported. There is a known limitation where files in a Microsoft Binder file cannot be extracted.

4.19 DASaveRandomTreeRecord

This function is called to extract a record in an archive file to disk. It is similar to DASaveTreeRecord, except that instead of reading the data for all nodes in the archive in a sequential order, this function will only read the data for the requested nodes from the archive. To use this function, you must first process the archive with Content Access or Search Export and save the Node Locator data for later use in this function.

```
DAERR DASaveRandomTreeRecord(
    VTHDOC          hDoc,
    SOTREENODELOCATOR sTreeNodeLocator,
    VTDWORD         dwSpecType,
    VTLPVOID        pSpec,
    VTDWORD         dwFlags)
```

Parameters

- hDoc: Identifier of open document. This hDoc must come from an archive document opened with DAOpenDocument with the flag DAOPENDOCUMENT_ARCHIVEONLYMODE set.
- sTreeNodeLocator: An SOTREENODELOCATOR structure which contains data identifying the desired node. This data should come from a previous conversion of the archive document using Content Access or Search Export.
- dwSpecType: Describes the contents of pSpec. Together, dwSpecType and pSpec describe the location of the source file to which the file will be extracted. Must be one of the following values:
 - IOTYPE_ANSIPATH: Windows only. pSpec points to a NULL-terminated full path name using the ANSI character set and FAT 8.3 (Win16) or NTFS (Win32 and Win64) filename conventions.
 - IOTYPE_REDIRECT: Specifies that redirected I/O will be used to save the file.
 - IOTYPE_UNICODEPATH: Windows only. pSpec points to a NULL-terminated full path name using the Unicode character set and NTFS (Win32 and Win64) file name conventions.
 - IOTYPE_UNIXPATH: X Windows on UNIX platforms only. pSpec points to a NULL-terminated full path name using the system default character set and UNIX path conventions. Unicode paths can be accessed on UNIX platforms by using a UTF-8 encoded path with IOTYPE_UNIXPATH.
- pSpec: File location specification
- dwFlags: Currently not used. Should be set to 0.

Return Value

- DAERR_OK: Returned if DASaveTreeRecord was successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.19.1 DATREENODELOCATOR

```
typedef struct DATREENODELOCATORtag
{
    VTDWORD dwSize; /* size of this structure */
    VTDWORD dwSpecialFlag; /* special flags coming from CA or SX */
    VTDWORD dwData1; /* dwData1 coming from CA or SX */
    VTDWORD dwData2; /* dwData2 coming from CA or SX */
}SCCDATREENODELOCATOR, *PSCCDATREENODELOCATOR;
```

4.19.2 SCCCA_TREENODELOCATOR: Tree Node Locator

This content type contains information to be used in the SOTREENODELOCATOR structure, which is used by [DAOpenRandomTreeRecord](#) and [DASaveRandomTreeRecord](#).

4.20 DACloseTreeRecord

This function is called to close an open record file handle.

Search Export Only: Search Export's default behavior is to automatically open and process the contents of an archive. Use DACloseTreeRecord and SCCOPT_XML_SEARCHML_FLAGS to change the default behavior if discrete processing of each document in an archive is desired.

```
DAERR DACloseTreeRecord(
    VTHDOC hDoc);
```

Parameters

- hDoc: Identifier of open record document.

Return Value

- DAERR_OK: Returned if DACloseTreeRecord was successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.21 DASETSTATCALLBACK

This function sets up a callback that the technology periodically calls to verify the file is still being processed. The customer can use this with a monitoring process to help identify files that may be hung. Because this function is called more frequently than other callbacks, it is implemented as a separate function.

Use of the Status Callback Function

An application's status callback function will be called periodically by Oracle Outside In to provide a status message. Currently, the only status message defined is OIT_STATUS_WORKING, which provides a "sign of life" that can be used during unusually long processing operations to verify that Oracle Outside In has not stopped working. If the application decides that it would not like to continue processing the current document, it may use the return value from this function to tell Oracle Outside In to abort.

The status callback function has two return values defined:

- OIT_STATUS_CONTINUE: Tells Oracle Outside In to continue processing the current document.

- OIT_STATUS_ABORT: Tells Oracle Outside In to stop processing the current document.

The following is an example of a minimal status callback function.

```

VTDWORD MyStatusCallback( VTHANDLE hUnique, VTDWORD dwID, VTSYSVAL
pCallbackData, VTSYSVAL pAppData)
{
    if(dwID == OIT_STATUS_WORKING)
    {
        if( checkNeedToAbort( pAppData ) )
            return (OIT_STATUS_ABORT);
    }

    return (OIT_STATUS_CONTINUE);
}

```

Prototype

```
DAERR DASetStatCallback(DASTATCALLBACKFN pCallback)
```

Parameters

- pCallback: Pointer to the callback function.
- dwID: Handle that indicates the callback status.
 - OIT_STATUS_WORKING
 - OIT_STATUS_CONTINUE
 - OIT_STATUS_ABORT
- pCallbackData: Currently always NULL

Return Values

- DAERR_OK: If successful. Otherwise, one of the other DAERR_ values in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

4.22 DASetFileAccessCallback

This function sets up a callback that the technology will call into to request information required to open an input file. This information may be the password of the file or a support file location.

Use of the File Access Callback

When the technology encounters a file that requires additional information to access its contents, the application's callback function will be called for this information. Currently, only two different forms of information will be requested: the password of a document, or the file used by Lotus Notes to authenticate the user information.

The status callback function has two return values defined:

- SCCERR_OK: Tells Oracle Outside In that the requested information is provided.
- SCCERR_CANCEL: Tells Oracle Outside In that the requested information is not available.

This function will be repeatedly called if the information provided is not valid (such as the wrong password). It is the responsibility of the application to provide the correct information or return SCCERR_CANCEL.

Prototype

```
DAERR DASetFileAccessCallback (DAFILEACCESSCALLBACKFN pCallback);
```

Parameters

- pCallback: Pointer to the callback function.

Return Values

- DAERR_OK: If successful. Otherwise, one of the other DAERR_ values defined in sccda.h or one of the SCCERR_ values in sccerr.h is returned.

The callback function should be of type DAFILEACCESSCALLBACKFN. This function has the following signature:

```
typedef VTDWORD (* DAFILEACCESSCALLBACKFN)(VTDWORD dwID, VTSYSVAL pRequestData,
VTSYSVAL pReturnData, VTDWORD dwReturnDataSize);
```

- dwID: ID of information requested:
 - OIT_FILEACCESS_PASSWORD: Requesting the password of the file
 - OIT_FILEACCESS_NOTESID: Requesting the Notes ID file location
- pRequestData – Information about the file.


```
typedef struct {
    VTDWORD    dwSize;           /* size of this structure */
    VTWORD     wFIId;           /* FI id of reference file */
    VTDWORD    dwSpecType;      /* file spec type */
    VTVOID     *pSpec;          /* pointer to a file spec */
    VTDWORD    dwRootSpecType;  /* root file spec type */
    VTDWORD    dwAttemptNumber; /* The number of times the callback has
                                /* already been called for the currently
                                /* requested item of information */
} IOREQUESTDATA, * PIOREQUESTDATA;
```
- pReturnData: Pointer to the buffer to hold the requested information – for OIT_FILEACCESS_PASSWORD and OIT_FILEACCESS_NOTESID, the buffer is an array of WORD characters.
- dwReturnDataSize: Size of the return buffer.

Note: Not all formats that use passwords are supported. Only Microsoft Office binary (97-2003) and Microsoft Office 2007, Lotus NSF, PDF (with RC4 encryption), Zip (with AES 128 & 256 bit, ZipCrypto) are currently supported.

Export Functions

This chapter outlines the basic functions used to initiate the conversion of documents using the product API.

5.1 General Functions

The following functions are general functions used in most products.

This section covers the following functions:

- [Section 5.1.1, "EXOpenExport"](#)
- [Section 5.1.2, "EXCloseExport"](#)
- [Section 5.1.3, "EXRunExport"](#)
- [Section 5.1.4, "EXExportStatus"](#)

5.1.1 EXOpenExport

This function is used to initiate the export process for a file that has been opened by DAOpenDocument. If EXOpenExport succeeds, EXCloseExport must be called regardless of any other API calls.

Prototype

```
SCCERR EXOpenExport (
    VTHDOC      hDoc,
    VTDWORD     dwOutputId,
    VTDWORD     dwSpecType,
    VTLPVOID    pSpec,
    VTDWORD     dwFlags,
    VTSYSPARAM  dwReserved,
    VTLPVOID    pCallbackFunc,
    VTSYSPARAM  dwCallbackData,
    VTLPHEXPORT phExport);
```

phExport is *not* a file handle.

Parameters

- hDoc: A handle that identifies the source file, created by DAOpenDocument. Knowledge of this should only affect OEMs under the most unusual of circumstances.
- dwOutputId: File ID of the desired format of the output file. This value must be set to the following values:

- FI_SEARCHML_LATEST
- FI_PAGEML
- FI_SEARCHTEXT
- FI_SEARCHHTML
- dwSpecType: Describes the contents of pSpec. Together, dwSpecType and pSpec describe the location of the initial output file. Must be one of the following values:
 - IOTYPE_ANSIPATH: Windows only. pSpec points to a NULL-terminated full path name using the ANSI character set and FAT 8.3 (Win16) or NTFS (Win32 and Win64) file name conventions.
 - IOTYPE_UNICODEPATH: Windows only. pSpec points to a NULL-terminated full path name using the Unicode character set and NTFS file name conventions.

Note: If you are using IOTYPE_UNICODEPATH as a file spec type, if the calling application is providing an export callback function, you should set the option SCCOPT_EX_UNICODECALLBACKSTR to TRUE. Refer to the documentation on callbacks such as EX_CALLBACK_ID_CREATENEWFILE and the EXURLFILEIOCALLBACKDATAW structure for details

- IOTYPE_UNIXPATH: X Windows on UNIX platforms only. pSpec points to a NULL-terminated full path name using the system default character set and UNIX path conventions. Unicode paths can be accessed on UNIX platforms by using a UTF-8 encoded path with IOTYPE_UNIXPATH.
- IOTYPE_REDIRECT: All platforms. A pointer to a BASEIO structure filled in by your application. This must not be set to NULL or conversion fails.
- pSpec: Initial output file location specification. The form of this data depends on the value of the dwSpecType parameter (see above). Initial output file location specification. This is either a pointer to a buffer or NULL.
- dwFlags: Must be set by developer to 0.
- dwReserved: Reserved. Must be set by developer to 0.
- phExport: Pointer to a handle that receives a value uniquely identifying the document to the product routines. If the function fails, this value is set to VTHDOC_INVALID.

Return Values

- SCCERR_OK: If the open was successful. Otherwise, one of the other SCCERR_ values in scerr.h is returned.

5.1.2 EXCloseExport

This function is called to terminate the export process for a file.

Prototype

```
SCCERR EXCloseExport(  
    VTHEXPORT hExport);
```


Parameters

- `hExport`: Export handle for the document. Must be a handle returned by the `EXOpenExport` function.

Return Values

- `SCCERR_OK`: Returned if the close was successful. Otherwise, one of the other `SCCERR_` values in `scerr.h` is returned.

5.1.3 EXRunExport

This function is called to run the export process.

Prototype

```
SCCERR EXRunExport(
    VTHEXPORT hExport);
```

Parameters

- `hExport`: Export handle for the document. Must be a handle returned by the `EXOpenExport` function.

Return Values

- `SCCERR_OK`: Returned if the export was successful. Otherwise, one of the other `SCCERR_` values in `scerr.h` is returned.

5.1.4 EXExportStatus

This function is used to determine if there were conversion problems during an export. It can either return detailed information about sub-document failures or a structure that describes areas of a conversion that may not have high fidelity with the original document.

Prototype

```
SCCERR EXExportStatus(VTHEXPORT hExport, VTDWORD dwStatusType, VTLPOID pStatus)
```

Parameters

- `hExport`: Export handle for the document.
- `dwStatusType`: Specifies which status information should be filled in `pStatus`.
 - `EXSTATUS_SUBDOC` – fills in the `EXSUBDOCSTATUS` structure (only implemented in Search Export and XML Export)
 - `EXSTATUS_INFORMATION` - fills in the `EXSTATUSINFORMATION` structure.
- `pStatus`: Either a pointer to a `EXSUBDOCSTATUS` or `EXSTATUSINFORMATION` data structure depending on the value of `dwStatusType`.

Return Values

`SCCERR_OK`: Returned if there were no problems. Otherwise, one of the other `SCCERR_` values in `scerr.h` is returned.

EXSUBDOCSTATUS Structure

The `EXSUBDOCSTATUS` structure is defined as follows:

```
typedef struct EXSUBDOCSTATUStag
{
    VTDWORD dwSize;          /* size of this structure */
    VTDWORD dwSucceeded;    /* number of sub documents that were converted */
    VTDWORD dwFailed;      /* number of sub documents that were not converted */
} EXSUBDOCSTATUS;
```

EXSTATUSINFORMATION Structure

The EXSTATUSINFORMATION structure is defined as follows:

```
typedef struct EXSTATUSINFORMATIONtag
{
    VTDWORD dwVersion;      /* version of this structure, currently
    EXSTATUSVERSION1      */
    VTBOOL bMissingMap;    /* a PDF text run was missing the toUnicode table
    */
    VTBOOL bVerticalText;  /* a vertical text run was present */
    VTBOOL bTextEffects;   /* unsupported text effects applied (i.e.Word
    Art)*/
    VTBOOL bUnsupportedCompression; /* a graphic had an unsupported compression */
    VTBOOL bUnsupportedColorSpace; /* a graphic had an unsupported color space */
    VTBOOL bForms;        /* a sub documents had forms */
    VTBOOL bRightToLeftTables; /* a table had right to left columns */
    VTBOOL bEquations;    /* a file had equations*/
    VTBOOL bAliasedFont;  /* A font was missing, but a font alias was used
    */
    VTBOOL bMissingFont;  /* The desired font wasn't present on the system
    */
    VTBOOL bSubDocFailed; /* a sub document was not converted */
} EXSTATUSINFORMATION;

#define EXSTATUSVERSION1 0X0001
```

Note: When processing the main document, Search Export, HTML Export, and XML Export never use fonts, so bAliasedFont and bMissingFont will never report TRUE; however, when doing graphics conversions XML Export and HTML Export may use fonts, so bAliasedFont and bMissingFont may report TRUE.

Anywhere a file specification (dwSpecType and pSpec parameters) is passed to a function in the product, the developer may use Redirected IO to completely take over responsibility for the low level IO calls of that particular file. The source file and all output files can be redirected in this way.

Redirected IO allows the developer great flexibility in the storage of, and access to, converted documents. For example, documents may be stored on file systems not supported natively by the software, or in a unique directory tree structure determined by the type of file.

This chapter includes the following sections:

- [Section 6.1, "Using Redirected IO"](#)
- [Section 6.2, "Opening Files"](#)
- [Section 6.3, "IOClose"](#)
- [Section 6.4, "IORead"](#)
- [Section 6.5, "IOWrite"](#)
- [Section 6.6, "IOSeek"](#)
- [Section 6.7, "IOTell"](#)
- [Section 6.8, "IOGetInfo"](#)
- [Section 6.9, "IOSEEK64PROC / IOTELL64PROC"](#)

6.1 Using Redirected IO

A developer can redirect the IO for an input or output file by providing a data structure that contains pointers to custom IO routines for reading and writing. This data structure is passed in place of a typical file specification. The developer must set the dwSpecType parameter of the DAOpenDocument call to IOTYPE_REDIRECT when the DAOpenDocument call is sent.

When dwSpecType is set this way, the pSpec element must contain a pointer to a developer-defined data structure that begins with a BASEIO structure (defined in baseIO.H). The BASEIO structure contains pointers to the basic IO functions for the IO system such as Read, Seek, Tell, etc. The developer must initialize these function pointers to their own functions that perform IO tasks. Beyond the BASEIO element, the developer may place any data he or she likes.

For instance, a developer's structure may be similar to the following:

```
typedef struct MYFILEtag
```

```
{
    BASEIO    sBaseIO;        /* must be the first element */
    VTDWORD   dwMyInfo1;
    VTDWORD   dwMyInfo2;
    .
    .
    .
} MYFILE;
```

Because the `pSpec` passed is essentially the "file handle" used by the software, the developer can redirect the IO on a file-by-file basis while still exporting "regular" disk-based files.

The `BASEIO` structure is defined as follows:

```
typedef struct BASEIOtag
{
    IOCLOSEPROC pClose;
    IOREADPROC pRead;
    IOWRITEPROC pWrite;
    IOSEEKPROC pSeek;
    IOTELLPROC pTell;
    IOGETINFOPROC pGetInfo;
    IOOPENPROC pOpen; /* pOpen *MUST* be set to NULL. */
#ifdef NLM
    IOSEEK64PROC pSeek64;
    IOTELL64PROC pTell64;
#endif
    VTVOID *aDummy[3];
} BASEIO, * PBASEIO;
```

The developer must implement the Close, Read, Write, Seek, Tell and GetInfo routines. The Open routine must be set to NULL. The first parameter to each of these routines is called `hFile` and is of the type `HIOFILE`. `HIOFILE` is simply the `VTLPVOID` to your data structure that was passed in the `pSpec` parameter of the `DAOpenDocument` call.

The sample source code for a simple implementation of Redirected IO is in the `samples` directory. This sample redirects the technology's IO through the `fopen`, `fgetc`, `fseek`, `ftell` and `fclose` run-time library routines.

Important: Redirected IO does not cache the whole file. Seeks can occur throughout the file during the course of conversion. If the developer is implementing redirected IO on a slow or sequential link, it is the developer's responsibility to cache the file locally.

6.2 Opening Files

The developer does not see a call to `pOpen` when using redirected IO. When `IOTYPE_REDIRECT` is used, the structure passed in `pSpec` is defined to represent a file that is already open. The software can immediately call the `pRead`, `pSeek`, `pTell` and `pWrite` functions.

Files specified as using redirected IO must be open by the time they are handed off to the software.

6.3 IOClose

Closes the file identified by `hFile` and cleans up all memory associated with the file.

If you dynamically allocate your own file structures (MYFILE in the preceding discussion) it is required that the memory allocated be freed inside the call to IOClose or sometime thereafter.

Prototype

```
IOERR IOClose(  
    HIOFILE  hFile);
```

Parameters

- hFile: Identifies the file to be closed. Should be cast into a pointer to your data structure (MYFILE in the preceding discussion).

Return Values

- IOERR_OK: Close was successful.
- IOERR_UNKNOWN: Some error occurred on close.

6.4 IORead

Reads data from the current file position forward and resets the position to the byte after the last byte read.

Prototype

```
IOERR IORead(  
    HIOFILE  hFile,  
    VTBYTE   * pData,  
    VTDWORD  dwSize,  
    VTDWORD  * pCount);
```

Parameters

- hFile: Identifies the file to be read. Should be cast into a pointer to your data structure (MYFILE in the preceding discussion).
- pData: Points to the buffer into which the bytes should be read. Will be at least dwSize bytes big.
- dwSize: Number of bytes to read.
- pCount: Points to the number of bytes actually read by the function. This value is only valid if the return value is IOERR_OK.

Return Values

- IOERR_OK: Read was successful. pCount contains the number of bytes read and pData contains the bytes themselves.
- IOERR_EOF: Read failed because the file pointer was beyond the end of the file at the time of the read.
- IOERR_UNKNOWN: Read failed for some other reason.

6.5 IOWrite

Writes data from the current file position forward and resets the position to the byte after the last byte written.

Prototype

```
IOERR IOWrite(  
    HIOFILE      hFile,  
    VTBYTE      * pData,  
    VTDWORD      dwSize,  
    VTDWORD      * pCount);
```

Parameters

- **hFile:** Identifies the file where the data is to be written. Should be cast into a pointer to your data structure (MYFILE in the preceding discussion).
- **pData:** Points to the buffer from which the bytes should be written. It must be at least `dwSize` bytes big. It is good practice to treat the data passed in by `pData` as "read only." This helps prevent unexpected behavior elsewhere in the system.
- **dwSize:** Number of bytes to write.
- **pCount:** Points to the number of bytes actually written by the function. This value is only valid if the return value is `IOERR_OK`.

Return Values

- `IOERR_OK:` Write was successful, `pCount` contains the number of bytes written.
- `IOERR_UNKNOWN:` Write failed for some reason.

6.6 IOSeek

Moves the current file position.

Prototype

```
IOERR IOSeek(  
    HIOFILE      hFile,  
    VTWORD       wFrom,  
    VTLONG       lOffset);
```

Parameters

- **hFile:** Identifies the file to be read. Should be cast into a pointer to your data structure (MYFILE in the preceding discussion).
- **wFrom:** One of the following values:
 - `IOSEEK_TOP:` Move the file position `lOffset` bytes from the top (beginning) of the file.
 - `IOSEEK_BOTTOM:` Move the file position `lOffset` bytes from the bottom (end) of the file.
 - `IOSEEK_CURRENT:` Move the file position `lOffset` bytes from the current file position.
- **lOffset:** Number of bytes to move the file pointer. A positive value moves the file pointer forward in the file and a negative value moves it backward. If a requested seek value would move the file pointer before the beginning of the file, the file pointer should remain unchanged and `IOERR_UNKNOWN` should be returned. Seeking past EOF is allowed. In that case `IOERR_OK` should be returned. `IOTell` would return the requested seek position and `IORead` should return `IOERR_EOF` and 0 bytes read.

Return Values

- IOERR_OK: Seek was successful.
- IOERR_UNKNOWN: Seek failed for some reason.

6.7 IOTell

Returns the current file position.

Prototype

```
IOERR IOTell(
    HIOFILE    hFile,
    VTDWORD    * pOffset);
```

Parameters

- hFile: Identifies the file to be read. Should be cast into a pointer to your data structure (MYFILE in the preceding discussion).
- pOffset: Points to the current file position returned by the function.

Return Values

- IOERR_OK: Tell was successful.
- IOERR_UNKNOWN: Tell failed for some reason.

6.8 IOGetInfo

Returns information about an open file.

Prototype

```
IOERR IOGetInfo(
    HIOFILE    hFile,
    VTDWORD    dwInfoId,
    VTVOID     * pInfo);
```

Parameters

- hFile: Identifies the file to be read. Should be cast into a pointer to your data structure (MYFILE in the previous discussion).
- dwInfoId: One of the following values:
 - IOGETINFO_FILENAME: pInfo points to a string that should be filled with the base file name (no path) of the open file (for example TEST.DOC). If you do not know the file name, return IOERR_UNKNOWN. Certain file types (such as DataEase) must know the original file name in order to open secondary files required to correctly view the original file. If you return IOERR_UNKNOWN, these file types do not convert. See [Section 6.8.1, "IOGENSECONDARY and IOGENSECONDARYW Structures."](#)
 - IOGETINFO_PATHNAME: pInfo points to a string that should be filled with the fully qualified path name (including the file name) of the open file. For example, C:\MYDIR\TEST.DOC. If you do not know the path name, return IOERR_UNKNOWN.
 - IOGETINFO_PATHTYPE: pInfo points to a DWORD that should be filled with the IOTYPE of the path returned by IOGETINFO_PATHNAME. For instance,

if you return a DOS path name in the Unicode character set, you should return `IOTYPE_UNICODEPATH`. Even if redirected IO is in use, this should not be set to `IOTYPE_REDIRECT`. The value should reflect the style of path to be returned or any other values detailed in [Section 5.1.1, "EXOpenExport."](#)

- `IOGETINFO_ISOLE2STORAGE`: Must return `IOERR_FALSE`. `pInfo` is not used.
- `IOGETINFO_GENSECONDARY`: `pInfo` points to a structure of type `IOGENSECONDARY`. Some file types require supporting files to be opened. These supporting files may contain formatting information or extra data. When using HTML Export, templates may link to other templates, and the paths to those templates must be resolved. Correct handling of `IOGETINFO_GENSECONDARY` is critical to the operation of the Oracle Outside In technology. For a list of these file types, see [Section 6.8.2, "File Types That Cause IOGETINFO_GENSECONDARY."](#)

Because the developer is in total control of the IO for the primary file, the technology does not know how to generate a path to these secondary files or even if the secondary files are accessible through the regular file system. The `IOGETINFO_GENSECONDARY` call gives the developer a chance to resolve this problem by generating a new IO specification for the secondary file in question. The developer gets just the base file name (often embedded in the original document or generated from the primary file's name) of the secondary file.

The developer may either use one of the standard Oracle Outside In IO types or totally redirect the IO for the secondary file, as well. For more details, see [Section 6.8.1, "IOGENSECONDARY and IOGENSECONDARYW Structures."](#)

- `IOGETINFO_SUBDOC_SPEC`: This message should be handled only if the currently open file is an archive and a particular item within the archive is intended to be specified as the input file in a call to `DAOpenDocument`. In this case, `pInfo` points to a single-byte character string that should be filled with the subdocument specification of an item within the open file. For example, `item.2` specifies item 2 within the archive file. When specifying a subdocument specification, return `IOERR_OK`. Any other return values cause the results of this message to be ignored.
- `IOGETINFO_64BITIO`: For redirected I/O that wishes to use 64-bit seek/tell functions, your `IOGetInfo` function must respond `IOERR_TRUE` to this `dwInfoId`. In addition, the `pSeek64/pTell64` items in the `baseio` structure must be valid pointers to the proper function types.
- `IOGETINFO_DPATHNAME`: `pInfo` points to a structure of type `DPATHNAME`, which should be filled with the fully qualified path name (including the file name) of the open file, for example, `C:\MYDIR\TEST.DOC`. If you do not know the path name, return `IOERR_UNKNOWN`. The `dwPathLen` element contains the size of the buffer pointed to by the `pPath` element. If the buffer size is too small to contain the full path, modify `dwPathLen` to be the correct size of the buffer required to hold the path name in its `IOTYPE` character width including the `NULL` terminator and return `IOERR_INSUFFICIENTBUFFER`.

The following is a C data structure defined in `SCCIO.H`:

```
typedef struct DPATHNAMEtag
{
    VTDWORD dwPathLen;
    VTVOID *pPath;
```



```
} DPATHNAME, * PDPATHNAME;
```

Parameters

dwPathLen: Will be set to the number of bytes in the buffer pointed to by pPath. If the size of the buffer is insufficient, reset this element to the number of bytes required and return IOERR_INSUFFICIENTBUFFER.

pPath: Points to the buffer to be filled with the path name.

- IOGETINFO_GENSECONDARYDP: pInfo points to a structure of type IOGENSECONDARYDP. The dwSpecLen element contains the size of the buffer pointed to by the pSpec element. If the buffer size is too small to contain the spec, modify dwSpecLen to be the correct size of the buffer required to hold the path in its IOTYPE character width including the NULL terminator and return IOERR_INSUFFICIENTBUFFER.

The following is a C data structure defined in SCCIO.H:

```
typedef struct IOGENSECONDARYDPtag
{
    VTDWORD          dwSize;
    VTVOID *         pFileName;
    VTDWORD          dwSpecType;
    VTVOID *         pSpec;
    VTDWORD          dwSpecLen;
    VTDWORD          dwOpenFlags;
} IOGENSECONDARYDP, * PIOGENSECONDARYDP;
```

Parameters

dwSize: Will be set to sizeof (IOGENSECONDARYDP)

pFileName: A pointer to a string representing the file name of the secondary file that the technology requires. It is usually a name stored in the primary file (such as MYSTYLE.STY for a Word for DOS file) or a name generated from the primary file name. The primary file for a DataEase database has a .dba extension. The secondary name is the same file name but with a .dbm extension.

dwSpecType: The developer must fill this with the IOSPEC for the secondary file.

pSpec: On entry, this pointer points to an array of bytes or may be NULL (see dwSpecLen below). If the dwSpecType is set a regular IOTYPE such as IOTYPE_ANSIPATH, the developer may fill this array with the path name or structure required for that IOTYPE. If the developer is redirecting access to the secondary file, then dwSpecType will be IOTYPE_REDIRECT and the developer should replace pSpec with a pointer to a developer-defined structure that begins with the BASEIO structure (see [Section 6.1, "Using Redirected IO"](#)).

The file is supposed to be opened by the OEM's redirected IO code by the time they return the BASEIO struct. This is because the pOpen routine in the BASEIO struct is supposed to be NULL.

dwSpecLen: On entry, this is set to the size of the pSpec buffer. If the size of the buffer is insufficient, replace the value with the number of bytes required and return IOERR_INSUFFICIENTBUFFER.

dwOpenFlags: Set by the technology. A set of bit flags describing how the secondary file should be opened. Multiple flags may be used by bitwise OR-ing them together. The following flags are currently used:

- IOOPEN_READ: The secondary file should be opened for read.
- IOOPEN_WRITE: The secondary file should be opened for write. If the specified file already exists, its contents are erased when this flag is set.
- IOOPEN_CREATE: The secondary file should be created (if it does not already exist) and opened for write.

Any other value should return IOERR_BADINFOID.

- pInfo: The size of the pInfo buffer depends on the **dwInfoId** selected. For IOGETINFO_FILENAME and IOGETINFO_PATHNAME, the buffer is of size MAX_PATH characters (each character is either one byte or two, depending on PATHTYPE). The IOGETINFO_PATHTYPE buffer is the size of a VTDWORD.

Return Values

- IOERR_OK: GetInfo was successful.
- IOERR_TRUE: Affirmative response from a true or false GetInfo.
- IOERR_FALSE: Negative response from a true or false GetInfo.
- IOERR_BADINFOID: dwInfoId can not be handled by this file type.
- IOERR_INVALIDSPEC: The file spec is bad for this type.
- IOERR_UNKNOWN: GetInfo failed for some other reason.

6.8.1 IOGENSECONDARY and IOGENSECONDARYW Structures

These structures are passed to the developer through the IOGetInfo function. They allow the developer to tell the technology where a secondary file, needed by the conversion process, is located.

The SpecType of the original file determines which of these two structures is used. If the SpecType is IOTYPE_UNICODEPATH, IOGENSECONDARYW is used. pFileName points to a Unicode string terminated with a NULL WORD. For all other SpecTypes, IOGENSECONDARY is used and pFileName points to a string terminated with a NULL BYTE.

When using HTML Export, consider the situation where the software must access a secondary template file. In that case, the SpecType of the original template specified by the option SCCOPT_EX_TEMPLATE determines which of the two structures is used.

The following is a C data structure defined in SCCIO.H:

```
typedef struct
{
    VTDWORD      dwSize;
    VTLPBYTE     pFileName;
    VTDWORD      dwSpecType;
    VTLPVOID     pSpec;
    VTDWORD      dwOpenFlags
} IOGENSECONDARY, * PIOGENSECONDARY;

typedef struct
{
    VTDWORD      dwSize;
    VTLPWORD     pFileName;
```

```

    VTDWORD    dwSpecType;
    VTLPVOID   pSpec;
    VTDWORD    dwOpenFlags
} IOGENSECONDARYW, * PIOGENSECONDARYW;

```

Parameters

- **dwSize:** Will be set to `sizeof (IOGENSECONDARY)` or `sizeof (IOGENSECONDARYW)` (both of these values are the same).
- **pFileName:** A pointer to a string representing the file name of the secondary file that the technology requires. It is usually a name stored in the primary file (such as `MYSTYLE.STY` for a Word for DOS file) or a name generated from the primary file name. The primary file for a DataEase database has a `.dba` extension. The secondary name is the same file name but with a `.dbm` extension.
- **dwSpecType:** The developer must fill this with the `IOSPEC` for the secondary file.
- **pSpec:** On entry, this pointer points to an array of 1024 bytes. If the `dwSpecType` is set a regular `IOTYPE` such as `IOTYPE_ANSIPATH`, the developer may fill this array with the path name or structure required for that `IOTYPE`. If the developer is redirecting access to the secondary file, then `dwSpecType` will be `IOTYPE_REDIRECT` and the developer should replace `pSpec` with a pointer to a developer-defined structure that begins with the `BASEIO` structure (see [Section 6.1, "Using Redirected IO"](#)).

The file is supposed to be opened by the OEM's redirected IO code by the time they return the `BASEIO` struct. This is because the `pOpen` routine in the `BASEIO` struct is supposed to be `NULL`.

- **dwOpenFlags:** Set by the technology. A set of bit flags describing how the secondary file should be opened. Multiple flags may be used by bitwise OR-ing them together. The following flags are currently used:
 - `IOOPEN_READ`: The secondary file should be opened for read.
 - `IOOPEN_WRITE`: The secondary file should be opened for write. If the specified file already exists, its contents are erased when this flag is set.
 - `IOOPEN_CREATE`: The secondary file should be created (if it does not already exist) and opened for write.

6.8.2 File Types That Cause `IOGETINFO_GENSECONDARY`

The following file types cause `IOGETINFO_GENSECONDARY`:

- **Microsoft Word for DOS Versions 4, 5 and 6:** Used to open and read the style sheet file associated with the document. The filter degrades if the style sheet is not present.
- **Harvard Graphics DOS 3.x:** Used to open and read the individual slides within `ScreenShow` and `palette` files. Files with the extension `.ch3` are individual graphics or slides that can be opened using no secondary files. Files with the extension `.sy3` are `ScreenShows` that reference a list of `.ch3` files via the secondary file mechanism. There is also an optional `palette` file that can be referenced from a `.ch3` file, but the filter degrades if the `palette` file is not present.
- **R:Base:** Used to open and read required schema file. The `R:Base` data files are named `????2.rbf` but the data is useless without the schema file named `????1.rbf`. There is also a `????3.rbf` file associated with each database, but it is not used.

- Paradox 4.0 and Above: Used to open and read memo field data file. Paradox uses a separate file for all memo field data larger than 32 bytes.
- DataEase: Used to open and read the data file. DataEase databases include a .dba file that contains the schema (the file that the technology can identify as DataEase) and a .dbm file that contains the actual data.
- Templates (HTML Export): Any template that contains a {## link} will need to open the linked files. Additionally, when the root template is opened using redirected IO, each {## copy} macro in the template will result in a IOGETINFO_GENSECONDARY call, as well.

6.9 IOSEEK64PROC / IOTELL64PROC

These functions are for seek/tell using 64-bit offsets. These functions are not used by default. Rather, they are used if the IOGETINFO_64BITIO message returns IOERR_TRUE. This is so redirected I/O using strictly 32-bit I/O is unaffected.

6.9.1 IOSeek64

Moves the current file position.

Prototype

```
IOERR IOSeek64(  
HIOFILE hFile,  
VWORD wFrom,  
VTOFF_T offset);
```

Parameters

The parameter information is the same as for IOSeek(). However, the size of the VTOFF_T offset for IOSeek64() is 64-bit unlike the 32-bit offset in IOSeek().

6.9.2 IOTell64

Returns the current file position.

Prototype

```
IOERR IOTell64(  
HIOFILE hFile,  
VTOFF_T * pOffset);
```

Parameters

The parameter information is the same as for IOTell(). The only change is the use of a pointer to a 64-bit parameter for returning the offset.

Implementation Issues

This chapter covers some issues specific to using the Export products.

7.1 Running in 24x7 Environments

To ensure robust 24x7 performance in server applications embedding the different export products, it is strongly recommended that the technology be run in a process separate from the server's primary process.

The file filtering technology underlying the technology represents almost a quarter of a million lines of code. This code is expected to robustly deal with any stream of bytes, of any length (any file), in all cases. Oracle has dedicated, and continues to dedicate, significant effort into making this technology extremely robust. However, in real world situations, expect that some small number of malformed files may force the filters into unstable states. This generally results in either a memory exception (which can be trapped and recovered from gracefully), infinite loop or a wild pointer that causes the filter to write into memory that is part of the same process but does not belong to the filter. In the latter situation, this wild pointer condition cannot be trapped.

On the desktop this is not a significant problem since the number of files being dealt with is relatively small. In a 24x7 server environment, however, a wild pointer can be extremely disruptive to the server process and produce serious problems. The best solution for dealing with this problem is to run any application that reads complex file formats in a separate process. This solution protects the application from the susceptibility of filtering technology to the unknown quality of input files.

It must be stressed that files that lead to wild pointers or infinite loops occur very infrequently, usually as a result of a third-party conversion process or beta versions of applications. Oracle is committed to addressing these issues and to updating and expanding its testing tools and corpus of documents to proactively minimize this "garbage in, garbage out" problem.

7.2 Running in Multiple Threads or Processes

On certain platforms, export products may be run in a multithreaded or multiprocessing application. The thing to remember when doing so is that each thread must go through all the steps listed in [Chapter 1, "Introduction."](#)

Sample Applications

Each of the sample applications included in this SDK is designed to highlight a specific aspect of the technology's functionality. We ship built versions of these sample applications. The compiled executables should be in the root directory where the product is installed.

Note: To use Transformation Server, you will need to set the TSROOT variable to the location of the Transformation Server installed SDK. For example, for a Linux version of Transformation Server, you would set:
TSROOT=/user/jsmith/ts/ts_linux-x86-32_sdk/sdk.

The following copyright applies to all sample applications shipped with this product:

Copyright © Oracle 1993, 2013

All rights reserved.

You have a royalty-free right to use, modify, reproduce and distribute the Sample Applications (and/or any modified version) in any way you find useful, provided that you agree that Oracle has no warranty obligations or liability for any Sample Application files.

This chapter includes the following sections:

- [Section 8.1, "Building the Samples on a Windows System"](#)
- [Section 8.2, "An Overview of the Sample Applications"](#)
- [Section 8.3, "Accessing the SDK via a Java Wrapper"](#)

8.1 Building the Samples on a Windows System

Microsoft Visual Studio project files are provided for building each of the sample applications. For 32-bit versions of Windows, versions of the project files are provided for Visual Studio 6 (.dsp files) and Visual Studio 2005 (.vcproj files).

Because .vcproj files may not pick up the right compiler on their own, you need to make sure that you are building with the Win64 configuration in Visual Studio 2005. For 64-bit versions of Windows, only the Visual Studio 2005 versions are available.

The project files for the sample applications can be found in the \sdk\samplecode\win subdirectory of the Oracle Outside In SDK.

For specific information about building the sample applications on your UNIX OS, see [Chapter 3, "UNIX Implementation Details."](#)

8.2 An Overview of the Sample Applications

Here's a quick tour of the sample applications provided with this product. Not all of the sample applications are provided for both the Windows and UNIX platforms. See the heading of each application's subsection for clarification.

This section includes the following sample applications:

- [Section 8.2.1, "*sample"](#)
- [Section 8.2.3, "export \(Windows Only\)"](#)
- [Section 8.2.4, "exsimple"](#)
- [Section 8.2.5, "extract_archive"](#)
- [Section 8.2.6, "process_isolate"](#)
- [Section 8.2.7, "xxredir \(Search Export\)"](#)

8.2.1 *sample

The name of this sample application varies according to product (sxsample for Search Export).

The following is a basic implementation that uses the default settings for every option.

```
sxsample Inputfile Outputfile
```

This sample app provides a very simple demonstration of creating SearchML output.

8.2.2 batch_process_sx

batch_process_sx demonstrates running Search Export in a separate process on multiple input files. It also allows the timing of each run.

The application is executed from the command line and takes several possible parameters:

```
batch_process_sx -f inputfile -o outputfile or [-d inputdir -o outputdir]
[-i iterations] [-q[2]] [-b]
```

- -f specifies the name of a single input file.
- -d specifies the name of an input directory of files.
- -o specifies the name of an output file if -f is being used, or the name of an output directory if -d is being used.
- -i is an optional parameter specifying the number of iterations to perform.
- -q and -q2 diminish the output to the screen.
- -b increases the amount of content in the output including processing tags and sub-documents.

8.2.3 export (Windows Only)

This application was designed to facilitate the testing of the software and should not be assumed to be of commercial quality.

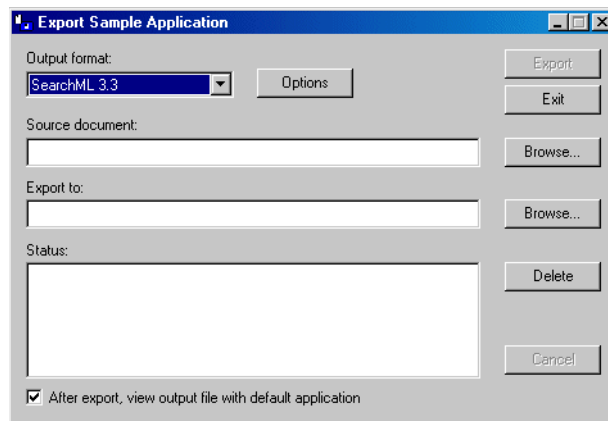
Important: No default options are set at initial runtime. The time the software is used, click the **Options** button and set the options. Failure to do this generates export errors.

The application allows the user to run a single source file. The user can choose the source file, an output file and set the various options.

8.2.3.1 The Export Main Window

The following figure shows the Main Window for the export application.

Figure 8–1 Export Main Window for Search Export



The Main Window is composed of several elements, discussed here.

- **Output Format menu:** This menu allows the user to select the type of output to generate. An entry for the format(s) you license will appear in this drop-down menu
- **Options button:** This opens up a new dialog with one or more tabs exposing the options for the selected product.
- **Source document field:** This is the document to be exported. Use the Browse button to pick the source file, or type in the path name.
- **'Export to' Field:** This is the initial resulting output file. Type in a file name or use the Browse button to choose a file. Other output files are named based on the one chosen here.
- **Delete button:** Clicking this button deletes all files generated by the last export, listed in the Status: field. This is useful when multiple output files are produced because the default naming rules do not overwrite an existing file. If you run Export over and over again with the same output file name, you can produce a large number of files. Clicking **Delete** before each export solves this problem.
- **'After Export, view output file with default application' checkbox:** If the export was successful, checking this box launches the initial output file in the application associated with the output flavor's default extension.
- **Export button:** Click this button to start the export process once you've determined the export settings.
- **Exit button:** Click to close the Export application.

8.2.4 exsimple

This simple command line driven program allows the user to run a single source file through the software. The user can choose the source file, an output file and set the various options.

To run the program, type:

```
exsimple in_file out_file config_file
```

- *in_file* is the input file to be converted
- *out_file* is the output location
- *config_file* is the configuration file that sets the conversion options. If no configuration file is specified, default.cfg in the current directory is used.

The configuration file is a text file used to set the conversion options. We recommend reading through the configuration file for more information about valid options and their values (use of invalid options results in exsimple not producing output).

Follow these instructions to set configurable options:

- Set the Output ID to FI_SEARCHHTML_LATEST before running the software. It is required and must not be omitted.

8.2.5 extract_archive

extract_archive demonstrates using the DATree API to extract all nodes in an archive.

The application is executed from the command line and takes two parameters, the name of the input file and the name of an output directory for the extracted files:

```
extract_archive input_file output_directory
```

8.2.6 process_isolate

process_isolate demonstrates running SX in a separate process. Due to the socket usage, there are different versions for Windows and Unix.

The application is executed from the command line and takes only one parameter, the name of the input file:

```
process_isolate_sample_unix -f input_file  
process_isolate_sample_win -f input_file
```

8.2.7 xxredir (Search Export)

This sample application is based on the exsimple sample application. It is designed to demonstrate how to use redirected IO and callbacks when using the software. It takes the same arguments and command line structure as exsimple and the same configuration files can be used. For details, see [Section 8.2.4, "exsimple."](#)

8.3 Accessing the SDK via a Java Wrapper

The ExJava Java wrapper, working in tandem with the exporter sample application, provides a working example of one method of interfacing with Oracle's C-based SDK products from a Java application. Export.jar is a Java API wrapper used by a Java application to control the exporter executable and set conversion options. exporter is a C-based executable which performs conversions using the modules in the Oracle Outside In SDK.

The exporter executable should be placed in the root directory of the Oracle Outside In SDK being used. If more than one Oracle Outside In SDK is being used, the contents of each SDK should be unpacked to the same root directory. Export.jar should be placed somewhere in your classpath.

On UNIX systems this sample application must be run from the directory containing the Oracle Outside In technology.

Java version 1.3.1 or higher is required to run this sample application.

This section covers the following topics:

- [Section 8.3.1, "The ExJava Wrapper API"](#)
- [Section 8.3.2, "The C-Based Exporter Application"](#)
- [Section 8.3.3, "Compiling the Executables"](#)
- [Section 8.3.4, "The ExportTest Sample Application"](#)
- [Section 8.3.5, "An Example Conversion Using the ExJava Wrapper"](#)

8.3.1 The ExJava Wrapper API

The JavaDocs documentation for the Java API is provided in the `/sdk/samplecode/ExJava/docs` directory. Conversion options are set using the `ExportProperties`.

Additionally, the appropriate `.cfg` file for the `ExportTest` sample application found in the `Examples/ExportTest` directory may provide further insight as to what properties are available and how they correspond to options and values for options.

The `Export.jar` and its source code can be found in the Java API directory. Place `Export.jar` somewhere in your classpath. In order to use the `ExportTest` sample application (which demonstrates how a Java application can use the ExJava API) without modifying your system configuration or the ExJava sample application, you should place the `Export.jar` file in the root directory of the Oracle Outside In SDK product you are using.

8.3.2 The C-Based Exporter Application

This is a standalone executable that runs out of process from the Java API. The Java API controls the conversion through command line parameters that are passed to the executable. After the conversion completes, the executable returns a conversion status code to the Java API. The command line parameters are base-64 encoded to allow for the use of Unicode encoded paths.

As the exporter executable is a C-based application, you will need to make sure the Java API can find the version of exporter appropriate for the platform you are using. Generally, and specifically for the purpose of using the `ExportTest` sample application, the correct executable should be copied to the root directory of the Oracle export SDK product you are using.

A compiled version of the C exporter program is included in the SDK with the rest of the Oracle Outside In binaries. The source for exporter is located in the `/sdk/samplecode/ExJava/exporter` directory.

The current implementation of ExJava may not produce an error if it cannot find the exporter application. This known issue may be corrected in a future version of ExJava.

8.3.3 Compiling the Executables

A Microsoft Visual Studio 6.0 project file and a UNIX makefile are provided in `/sdk/samplecode/ExJava/exporter/win` and `/sdk/samplecode/ExJava/exporter/unix`, respectively, so that you can modify the exporter executable or compile it for a platform other than those for which compiled versions of exporter are provided. If you unpacked the ExJava package into the root directory of one of Oracle's export SDK products, you should be able to use the Visual Studio Project and makefile as is. Otherwise, you will need to edit them in order to provide paths to the Oracle export SDK include and library files.

If you are compiling ExJava for use on the Solaris platform, make sure your `LD_LIBRARY_PATH` contains the Oracle Outside In SDK path before trying to build the exporter module.

8.3.4 The ExportTest Sample Application

ExportTest is an example of how a Java developer could use the ExJava wrapper to use one of the Oracle Outside In SDKs. The following is a list of the components that should be placed in the root directory of the Oracle Outside In SDK you are using in order to run this sample application:

1. `Export.jar` (from the Java API directory)
2. Exporter module for the platform you wish to use (located in the `/sdk/samplecode/ExJava/exporter/win` or `/sdk/samplecode/ExJava/exporter/unix` directory, depending on which platform you are using)
3. `sx.cfg` (also in `Examples/ExportTest` directory)
4. If you are running ExportTest on a UNIX system, make sure to edit the `.cfg` file so it reflects the correct name of the exporter module you renamed.
5. `ExportTest.jar` (also in `Examples/ExportTest` directory)
6. The appropriate batch file to run the ExportTest application (`ExportTest.bat` for Windows and `ExportTest.sh` for UNIX, both located in the `Examples/ExportTest` directory)

Once these files are properly copied, execute the batch file with the name/path of an input file to convert, the name for the base output file and the name of the configuration file to use for setting conversion options.

`ExportTest.jar` uses the contents of the configuration file to determine what option/value pairs it should use when doing the conversion. It is not necessary to use a configuration file when developing your own application if you so choose not to.

8.3.5 An Example Conversion Using the ExJava Wrapper

This is a simple outline of the steps for using the ExJava wrapper on a Windows system to convert a Word document called `MyWordDoc.Doc`. If you are using a UNIX system, see [Chapter 3, "UNIX Implementation Details"](#) for information about properly setting up your environment to use the Oracle Outside In SDK:

1. Edit the `.cfg` file and make sure `outputid` is set to the `FI*` value appropriate for the Oracle Outside In product you've licensed. Alter any other parameters in the `.cfg` file as needed then save the file.
2. Execute the following command. The sample command below assumes XML as the export type. Change this type accordingly:

```
ExportTest.bat myworddoc.doc output.xml sx.cfg
```

Copyrights and Licensing

This appendix provides a comprehensive overview of all copyright and licensing information for Oracle Outside In Search Export.

A.1 Oracle Outside In Search Export Licensing

The Programs (which include both the software and documentation) contain proprietary information; they are provided under a license agreement containing restrictions on use and disclosure and are also protected by copyright, patent, and other intellectual and industrial property laws. Reverse engineering, disassembly, or decompilation of the Programs, except to the extent required to obtain interoperability with other independently created software or as specified by law, is prohibited.

The information contained in this document is subject to change without notice. If you find any problems in the documentation, please report them to us in writing. This document is not warranted to be error-free. Except as may be expressly permitted in your license agreement for these Programs, no part of these Programs may be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose.

If the Programs are delivered to the United States Government or anyone licensing or using the Programs on behalf of the United States Government, the following notice is applicable:

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the Programs, including documentation and technical data, shall be subject to the licensing restrictions set forth in the applicable Oracle license agreement, and, to the extent applicable, the additional rights set forth in FAR 52.227-19, Commercial Computer Software--Restricted Rights (June 1987). Oracle USA, Inc., 500 Oracle Parkway, Redwood City, CA 94065.

The Programs are not intended for use in any nuclear, aviation, mass transit, medical, or other inherently dangerous applications. It shall be the licensee's responsibility to take all appropriate fail-safe, backup, redundancy and other measures to ensure the safe use of such applications if the Programs are used for such purposes, and we disclaim liability for any damages caused by such use of the Programs.

Oracle, JD Edwards, PeopleSoft, and Siebel are registered trademarks of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

The Programs may provide links to web sites and access to content, products, and services from third parties. Oracle is not responsible for the availability of, or any content provided on, third-party web sites. You bear all risks associated with the use of such content. If you choose to purchase any products or services from a third party, the relationship is directly between you and the third party. Oracle is not responsible for: (a) the quality of third-party products or services; or (b) fulfilling any of the terms of the agreement with the third party, including delivery of products or services and warranty obligations related to purchased products or services. Oracle is not responsible for any loss or damage of any sort that you may incur from dealing with any third party.

Portions relating to XServer copyright 1990, 1991 Network Computing Devices, 1987 Digital Equipment Corporation and the Massachusetts Institute of Technology.

Portions of this software are copyright © 1996-2002 The FreeType Project (www.freetype.org). All rights reserved.

Portions copyright 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002 by Cold Spring Harbor Laboratory. Funded under Grant P41-RR02188 by the National Institutes of Health.

Portions copyright 1996, 1997, 1998, 1999, 2000, 2001, 2002 by Boutell.Com, Inc.

Portions relating to GD2 format copyright 1999, 2000, 2001, 2002 Philip Warner.

Portions relating to PNG copyright 1999, 2000, 2001, 2002 Greg Roelofs.

Portions relating to PNG Copyright 1995-1996 Jean-loup Gailly and Mark Adler

Portions relating to PNG Copyright 1998, 1999 Glenn Randers-Pehrson, Tom Lane, Willem van Schaik, John Bowler, Kevin Bracey, Sam Bushell, Magnus Holmgren, Greg Roelofs, Tom Tanner, Andreas Dilger, Dave Martindale, Guy Eric Schalnat, Paul Schmidt, Tim Wegner

Portions relating to gdtf.c copyright 1999, 2000, 2001, 2002 John Ellson (ellson@graphviz.org).

Portions relating to gdft.c copyright 2001, 2002 John Ellson (ellson@graphviz.org).

Portions relating to JPEG and to color quantization copyright 2000, 2001, 2002, Doug Becker and copyright (C) 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, Thomas G. Lane. This software is based in part on the work of the Independent JPEG Group. See the file README-JPEG.TXT for more information.

Portions relating to WBMP copyright 2000, 2001, 2002 Maurice Szmurlo and Johan Van den Brande.

Portions relating to GIF Copyright 1987, by Steven A. Bennett.

Permission has been granted to copy, distribute and modify gd in any context without fee, including a commercial application, provided that this notice is present in user-accessible supporting documentation.

This does not affect your ownership of the derived work itself, and the intent is to assure proper credit for the authors of gd, not to interfere with your productive use of gd. If you have questions, ask. "Derived works" includes all programs that utilize the library. Credit must be given in user-accessible documentation.

This software is provided "AS IS." The copyright holders disclaim all warranties, either express or implied, including but not limited to implied warranties of merchantability and fitness for a particular purpose, with respect to this code and accompanying documentation.

Although their code does not appear in gd 2.0.4, the authors wish to thank David Koblas, David Rowley, and Hutchison Avenue Software Corporation for their prior contributions.

UnRAR - free utility for RAR archives

License for use and distribution of FREE portable version

The source code of UnRAR utility is freeware. This means:

1. All copyrights to RAR and the utility UnRAR are exclusively owned by the author - Alexander Roshal.
2. The UnRAR sources may be used in any software to handle RAR archives without limitations free of charge, but cannot be used to re-create the RAR compression algorithm, which is proprietary. Distribution of modified UnRAR sources in separate form or as a part of other software is permitted, provided that it is clearly stated in the documentation and source comments that the code may not be used to develop a RAR (WinRAR) compatible archiver.
3. The UnRAR utility may be freely distributed. No person or company may charge a fee for the distribution of UnRAR without written permission from the copyright holder.
4. THE RAR ARCHIVER AND THE UNRAR UTILITY ARE DISTRIBUTED "AS IS". NO WARRANTY OF ANY KIND IS EXPRESSED OR IMPLIED. YOU USE AT YOUR OWN RISK. THE AUTHOR WILL NOT BE LIABLE FOR DATA LOSS, DAMAGES, LOSS OF PROFITS OR ANY OTHER KIND OF LOSS WHILE USING OR MISUSING THIS SOFTWARE.
5. Installing and using the UnRAR utility signifies acceptance of these terms and conditions of the license.
6. If you don't agree with terms of the license you must remove UnRAR files from your storage devices and cease to use the utility.

JasPer License Version 2.0

Copyright (c) 2001-2006 Michael David Adams

Copyright (c) 1999-2000 Image Power, Inc.

Copyright (c) 1999-2000 The University of British Columbia

All rights reserved.

Permission is hereby granted, free of charge, to any person (the "User") obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

1. The above copyright notices and this permission notice (which includes the disclaimer below) shall be included in all copies or substantial portions of the Software.
2. The name of a copyright holder shall not be used to endorse or promote products derived from the Software without specific prior written permission.

THIS DISCLAIMER OF WARRANTY CONSTITUTES AN ESSENTIAL PART OF THIS LICENSE. NO USE OF THE SOFTWARE IS AUTHORIZED HEREUNDER EXCEPT UNDER THIS DISCLAIMER. THE SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED,

INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, OR ANY SPECIAL INDIRECT OR CONSEQUENTIAL DAMAGES, OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE. NO ASSURANCES ARE PROVIDED BY THE COPYRIGHT HOLDERS THAT THE SOFTWARE DOES NOT INFRINGE THE PATENT OR OTHER INTELLECTUAL PROPERTY RIGHTS OF ANY OTHER ENTITY. EACH COPYRIGHT HOLDER DISCLAIMS ANY LIABILITY TO THE USER FOR CLAIMS BROUGHT BY ANY OTHER ENTITY BASED ON INFRINGEMENT OF INTELLECTUAL PROPERTY RIGHTS OR OTHERWISE. AS A CONDITION TO EXERCISING THE RIGHTS GRANTED HEREUNDER, EACH USER HEREBY ASSUMES SOLE RESPONSIBILITY TO SECURE ANY OTHER INTELLECTUAL PROPERTY RIGHTS NEEDED, IF ANY. THE SOFTWARE IS NOT FAULT-TOLERANT AND IS NOT INTENDED FOR USE IN MISSION-CRITICAL SYSTEMS, SUCH AS THOSE USED IN THE OPERATION OF NUCLEAR FACILITIES, AIRCRAFT NAVIGATION OR COMMUNICATION SYSTEMS, AIR TRAFFIC CONTROL SYSTEMS, DIRECT LIFE SUPPORT MACHINES, OR WEAPONS SYSTEMS, IN WHICH THE FAILURE OF THE SOFTWARE OR SYSTEM COULD LEAD DIRECTLY TO DEATH, PERSONAL INJURY, OR SEVERE PHYSICAL OR ENVIRONMENTAL DAMAGE ("HIGH RISK ACTIVITIES"). THE COPYRIGHT HOLDERS SPECIFICALLY DISCLAIM ANY EXPRESS OR IMPLIED WARRANTY OF FITNESS FOR HIGH RISK ACTIVITIES.

Search Export Options

Options are parameters affecting the behavior of an export or transformation. This chapter presents both the C/C++ and SOAP options relevant to the Search Export product.

While default values are provided, users are encouraged to set all options for a number of reasons. In some cases, the default values were chosen to provide backwards compatibility. In other cases, the default values were chosen arbitrarily from a range of possibilities.

B.1 Search Export C/C++ Options

These options are available to the developer when using the export engine.

Options are set using the `DASetOption` call. It is recommended that developers familiarize themselves with all of the options available.

Options may be `Local`, in which case they only affect the handle for which they are set, or `Global`, in which case they automatically affect all handles associated with the `hDoc` and must be set before the call to `DAOpenDocument`.

B.1.1 Character Mapping

This section discusses character mapping options.

B.1.1.1 `SCCOPT_DEFAULTINPUTCHARSET`

This option is used in cases where Oracle Outside In cannot determine the character set used to encode the text of an input file. When all other means of determining the file's character set are exhausted, Oracle Outside In will assume that an input document is encoded in the character set specified by this option. This is most often used when reading plain-text files, but may also be used when reading HTML or PDF files. The possible character sets are listed in `charsets.h`.

When "extended test for text" is enabled (see [Section B.1.3.3, "SCCOPT_FIFLAGS"](#)), this option will still apply to plain-text input files that are not identified as EBCDIC or Unicode.

This option supersedes the `SCCOPT_FALLBACKFORMAT` option for selecting the character set assumed for plain-text files. For backwards compatibility, use of deprecated character-set -related values is still currently supported for `SCCOPT_FALLBACKFORMAT`, though internally such values will be translated into equivalent values for the `SCCOPT_DEFAULTINPUTCHARSET`. As a result, if an application were to set both options, the last such value set for either option will be the value that takes effect.

Handle Types

NULL, VTHDOC

Scope

Global

Data Type

VTDWORD

Default

- ANSI1252 on Windows and Latin-1 on UNIX.

Data

The data types are listed in charsets.h.

B.1.1.2 SCCOPT_UNMAPPABLECHAR

This option selects the character used when a character is not a valid Unicode character, or does not conform to the XML specification for valid characters. This option takes the Unicode value for the replacement character. If you are using the PageML output format, this option is only valid if the SCCEX_PAGEML_TEXTOUT flag is set in [SCCOPT_XML_PAGEML_FLAGS](#).

Handle Types

VTHDOC

Scope

Local

Data Type

VTWORD

Data

The Unicode value for the character to use.

Default

- 0xffff

B.1.2 Output

This section discusses output options.

B.1.2.1 SCCOPT_RENDERING_PREFER_OIT

This option is valid on 32-bit and 64-bit Linux (Red Hat and Suse) and Solaris Sparc platforms.

This option is only valid when PageML is the output format.

When this option is set to TRUE, the technology will attempt to use its internal graphics code to render fonts and graphics. When set to FALSE, the technology will render images using the operating system's native graphics subsystem (X11 on UNIX/Linux platforms). Note that this option only works when at least one of the

appropriate output solutions is present. For example, if the UNIX \$DISPLAY variable does not point to a valid X Server, but the OSGD and/or WV_GD modules required for the Oracle Outside In output solution exist, Oracle Outside In will default to the Oracle Outside In rendering code. The option will fail if neither of these output solutions is present.

It is important for the system to be able to locate useable fonts when this option is set to TRUE. Only TrueType fonts (*.ttf or *.ttc files) are currently supported. To ensure that the system can find them, make sure that the environment variable GDFONTPATH includes one or more paths to these files. If the variable GDFONTPATH can't be found, the current directory is used. If fonts are called for and cannot be found, Search Export will exit with an error. Also note that when copying Windows fonts to a UNIX system, the font extension for the files (*.ttf or *.ttc) must be lowercase, or they will not be detected during the search for available fonts. Oracle does not provide fonts with any Oracle Outside In product.

Handle Types

NULL, VTHDOC

Scope

Global

Data Type

VTBOOL

Data

One of the following values:

- TRUE: Use the technology's internal graphics rendering code to produce bitmap output files whenever possible.
- FALSE: Use the operating system's native graphics subsystem.

Default

FALSE

B.1.3 Input Handling

This section discusses input handling options.

B.1.3.1 SCCOPT_EXTRACTXMPMETADATA

Adobe's Extensible Metadata Platform (XMP) is a labeling technology that allows you to embed data about a file, known as metadata, into the file itself. This option enables the XMP feature, which does not interpret the XMP metadata, but passes it straight through without any interpretation. This option is independent of the other two "metadata" options. This option will be ignored if the SCCOPT_PARSEXMPMETADATA option is enabled.

- SCCEX_IND_SUPPRESSPROPERTIES will not affect XMP, so if you turn XMP on, but also set SuppressProperties, you will still get the XMP.
- SCCEX_METADATAONLY will not guarantee that XMP is produced.

Handle Types

VTHDOC

Scope

Local (was Global prior to release 8.2.2)

Data Type

VTBOOL

Data

- TRUE: This setting enables XMP extraction.
- FALSE: This setting disables XMP extraction.

Default

- FALSE

B.1.3.2 SCCOPT_FALLBACKFORMAT

This option controls how files are handled when their specific application type cannot be determined. This normally affects all plain-text files, because plain-text files are generally identified by process of elimination, for example, when a file isn't identified as having been created by a known application, it is treated as a plain-text file.

This option must be set for an hDoc before any subhandle has been created for that hDoc.

A number of values that were formerly allowed for this option have been deprecated. Specifically, the values that selected specific plain-text character sets are no longer to be used. Instead, applications should use the [SCCOPT_DEFAULTINPUTCHARSET](#) option for such functionality.

Handle Types

NULL, VTHDOC

Scope

Global

Data Type

VTDWORD

Data

The high VTWORD of this value is reserved and should be set to 0, and the low VTWORD must have one of the following values:

- FI_TEXT: Unidentified file types will be treated as text files.
- FI_NONE: Oracle Outside In will not attempt to process files whose type cannot be identified. This will include text files. When this option is selected, an attempt to process a file of unidentified type will cause Oracle Outside In to return an error value of DAERR_FILTERNOTAVAIL (or SCCERR_NOFILTER).

Default

- FI_TEXT

B.1.3.3 SCCOPT_FIFLAGS

This option affects how an input file's internal format (application type) is identified when the file is first opened by the Oracle Outside In technology. When the extended test flag is in effect, and an input file is identified as being either 7-bit ASCII, EBCDIC, or Unicode, the file's contents will be interpreted as such by the export process.

The extended test is optional because it requires extra processing and cannot guarantee complete accuracy (which would require the inspection of every single byte in a file to eliminate false positives.)

Handle Types

NULL, VTHDOC

Scope

Global

Data Type

VTDWORD

Data

One of the following values:

- **SCCUT_FI_NORMAL**: This is the default value. When this is set, standard file identification behavior occurs.
- **SCCUT_FI_EXTENDEDTEST**: If set, the File Identification code will run an extended test on all files that are not identified.

Default

- **SCCUT_FI_NORMAL**

B.1.3.4 SCCOPT_FORMATFLAGS

This option allows the developer to set flags that enable options that span multiple export products.

Handle Types

VTHDOC

Scope

Local

Data Type

VTDWORD

Data

- **SCCOPT_FLAGS_ALLISODATETIMES**: When this flag is set, all Date and Time values are converted to the ISO 8601 standard. This conversion can only be performed using dates that are stored as numeric data within the original file.
- **SCCOPT_FLAGS_STRICTFILEACCESS**: When an embedded file or URL can't be opened with the full path, OIT will sometimes try and open the referenced file from other locations, including the current directory. When this flag is set, it will

prevent OIT from trying to open the file from any location other than the fully qualified path or URL.

Default

0: All flags turned off

B.1.3.5 SCCOPT_SYSTEMFLAGS

This option controls a number of miscellaneous interactions between the developer and the Outside In Technology.

Handle Type

VTHDOC

Scope

Local

Data Type

VTDWORD

Data

- `SCCVW_SYSTEM_UNICODE`: This flag causes the strings in `SCCDATREENODE` to be returned in Unicode.

Default

0

B.1.3.6 SCCOPT_IGNORE_PASSWORD

This option can disable the password verification of files where the contents can be processed without validation of the password. If this option is not set, the filter should prompt for a password if it handles password-protected files.

As of Release 8.4.0, only the PST and MDB Filters support this option.

Scope

Global

Data Type

VTBOOL

Data

- `TRUE`: Ignore validation of the password
- `FALSE`: Prompt for the password

Default

FALSE

B.1.3.7 SCCOPT_LOTUSNOTESDIRECTORY

This option allows the developer to specify the location of a Lotus Notes or Domino installation for use by the NSF filter. A valid Lotus installation directory must contain the file `nnotes.dll`.

Note: Please see section 2.1.1 for NSF support on Win x86-32 or Win x86-64 or section 3.1.1 for NSF support on Linux x86-32 or Solaris Sparc 32.

Handle Types

NULL

Scope

Global

Data Type

VTLPBYTE

Data

A path to the Lotus Notes directory.

Default

If this option isn't set, then OIT will first attempt to load the Lotus library according to the operating system's PATH environment variable, and then attempt to find and load the Lotus library as indicated in HKEY_CLASSES_ROOT\Notes.Link.

B.1.3.8 SCCOPT_PARSEXMPMETADATA

Adobe's Extensible Metadata Platform (XMP) is a labeling technology that allows you to embed data about a file, known as metadata, into the file itself. This option enables parsing of the XMP data into normal OIT document properties. Enabling this option may cause the loss of some regular data in premium graphics filters (such as Postscript), but won't affect most formats (such as PDF).

Handle Types

VTHDOC

Scope

Local

Data Type

VTBOOL

Data

- TRUE: This setting enables parsing XMP.
- FALSE: This setting disables parsing XMP.

Default

FALSE

B.1.3.9 SCCOPT_PDF_FILTER_REORDER_BIDI

This option controls whether or not the PDF filter will attempt to reorder bidirectional text runs so that the output is in standard logical order as used by the Unicode 2.0 and later specification. This additional processing will result in slower filter performance according to the amount of bidirectional data in the file.

Handle Types

VTHDOC, NULL

Scope

Global

Data Type

VTDWORD

Data

- SCCUT_FILTER_STANDARD_BIDI
- SCCUT_FILTER_REORDERED_BIDI

Default

SCCUT_FILTER_STANDARD_BIDI

B.1.3.10 SCCOPT_PROCESS_OLE_EMBEDDINGS

Microsoft Powerpoint versions from 1997 through 2003 had the capability to embed OLE documents in the Powerpoint files. This option controls which embeddings are to be processed as native (OLE) documents and which are processed using the alternate graphic.

Note: The Microsoft Powerpoint application sometimes does embed known Microsoft OLE embeddings (such as Visio, Project) as an "Unknown" type. To process these embeddings, the SCCOPT_PROCESS_OLEEMBED_ALL option is required. Post Office-2003 products such as Office 2007 embeddings also fall into this category.

Handle Types

VTHDOC, NULL

Scope

Global

Data Type

VTWORD

Data

- SCCOPT_PROCESS_OLEEMBED_ALL : Process all embeddings in the file
- SCCOPT_PROCESS_OLEEMBED_NONE : Process none of the embeddings in the file
- SCCOPT_PROCESS_OLEEMBED_STANDARD (default) : Process embeddings that are known standard embeddings. These include Office 2003 versions of Word, Excel, Visio etc.

Default

SCCOPT_PROCESS_OLEEMBED_STANDARD

B.1.3.11 SCCOPT_TIMEZONE

This option allows the user to define an offset to GMT that will be applied during date formatting, allowing date values to be displayed in a selectable time zone. This option affects the formatting of numbers that have been defined as date values. This option will not affect dates that are stored as text.

Note: This option does not apply for spreadsheet files.

Handle Types

NULL, VTHDOC

Scope

Global

Data Type

VTLONG

Data

Integer parameter from -96 to 96, representing 15-minute offsets from GMT. To query the operating system for the time zone set on the machine, specify `SCC_TIMEZONE_USENATIVE`.

Default

- 0: GMT time

B.1.3.12 SCCOPT_HTML_COND_COMMENT_MODE

Some HTML includes a special type of comment that will be read by particular versions of browsers or other products. This option allows you to control which of those comments are included in the output.

Handle Type

VTHDOC

Scope

Local

Data Type

VTDWORD

Data

- One or more of the following values OR-ed together:
- `HTML_COND_COMMENT_NONE`: Don't output any conditional comments.
Note: setting any other flag will negate this.
- `HTML_COND_COMMENT_IE5`: include the IE 5 comments
- `HTML_COND_COMMENT_IE6`: include the IE 6 comments
- `HTML_COND_COMMENT_IE7`: include the IE 7 comments
- `HTML_COND_COMMENT_IE8`: include the IE 8 comments

- HTML_COND_COMMENT_IE9: include the IE 9 comments
- HTML_COND_COMMENT_ALL: include all conditional comments including the versions listed above and any other versions that might be in the HTML.

B.1.3.13 SCCOPT_PDF_FILTER_DROPHYPHENS

This option controls whether or not the PDF filter will drop hyphens at the end of a line. Since most PDF-generating tools create them as generic dashes, it's impossible for Outside In to know if the hyphen is a syllable hyphen or part of a hyphenated word. When this option is set to TRUE, all hyphens at the end of lines will be dropped from the extracted text.

Note: When this option is TRUE, the character counts for the extracted text may not match the counts used for rendering where the hyphens are required for rendering. This will affect annotations in rendering APIs.

Handle Types

VTHDOC

Scope

Local

Data Type

VTBOOL

Data

- TRUE: This setting drops hyphens from the end of all lines.
- FALSE: This setting retains hyphens at the end of all lines.

Default

FALSE

B.1.3.14 SCCOPT_ARCFULLPATH

In the Viewer and rendering products, this option tells the archive display engine to show the full path to a node in the szNode field in response to a SCCVW_GETTREENODE message. It also causes the name fields in DAGetTreeRecord and DAGetObjectInfo to contain the full path instead of just the archive node name.

Data Type

VTBOOL

Data

- TRUE: Display the full path.
- FALSE: Do not display the path.

Default

FALSE

B.1.4 Compression

This section discusses compression options.

B.1.4.1 SCCOPT_FILTERLZW

This option can disable access to any files using Lempel-Ziv-Welch (LZW) compression, such as .GIF files, .ZIP files or self-extracting archive (.EXE) files containing "shrunk" files. Attempts to read such files when this option is enabled will fail and return the error SCCERR_UNSUPPORTEDCOMPRESSION.

The following is a list of file types affected when this option is disabled:

- GIF files
- TIF files using LZW compression
- PDF files that use internal LZW compression
- TAZ and TAR archives containing files that are identified as FI_UNIXCOMP
- ZIP and self-extracting archive (.EXE) files containing "shrunk" files
- Postscript files using LZW compression

Although this option can disable access to files in ZIP or EXE archives stored using LZW compression, any files in such archives that were stored using any other form of compression will still be accessible.

Handle Types

VTHDOC, HEXPORT

Scope

Local

Data Type

VTDWORD

Data

- SCCVW_FILTER_LZW_ENABLED: LZW compressed files will be read normally.
- SCCVW_FILTER_LZW_DISABLED: LZW compressed files will not be read.

Default

SCCVW_FILTER_LZW_ENABLED

B.1.5 XML

This section discusses XML options.

B.1.5.1 SCCOPT_ENABLEALLSUBOBJECTS

Oracle Outside In has an internal flag that is used to optimize several of the input filters for searching. One of the side effects of this optimization is that many embedded bitmaps, including Progressive JPEG, aren't output by the filter. SCCOPT_ENABLEALLSUBOBJECTS can override this internal optimization.

Handle Types

VTHDOC

Scope

Local

Data Type

VTDWORD

Data

One of the following values:

- `SCCVW_FILTER_ENABLEALLSUBOBJECTS`: Override the optimizations.
- `SCCVW_FILTER_NORMALSUBOBJECTS`: Allow the optimizations.

Default`SCCVW_FILTER_NORMALSUBOBJECTS`**B.1.5.2 SCCOPT_XML_DEF_METHOD**

This option determines whether Search Export will reference a SearchML or PageML schema, DTD, or no reference when generating output. This option is not valid when SearchText or SearchHTML is the output format.

Handle Types

VTHDOC

Scope

Local

Data Type

VTDWORD

Data

One of the following values:

- `SCCEX_XML_XDM_DTD`: Document Type Definition (DTD)
- `SCCEX_XML_XDM_XSD`: Extensible Schema Definition
- `SCCEX_XML_XDM_NONE`: No XML definition reference

Default`SCCEX_XML_XDM_NONE`**B.1.5.3 SCCOPT_XML_DEF_REFERENCE**

This option allows the developer to set a particular file as the XML definition reference.

If the [SCCOPT_XML_DEF_METHOD](#) option is set to `SCCEX_XML_XDM_XSD` or `SCCEX_XML_XDM_DTD`, the value of this option will be used to reference the schema or DTD, respectively.

Handle Types

VTHDOC

Scope

Local

Data Type

Size (in bytes) of the data being passed, including a terminating NULL.

Data

The size of an array that holds WORD-sized characters terminated with a WORD-sized NULL (a UCS-2 string). The size passed is the total number of bytes that this UCS-2 string comprises. It includes in its size the bytes occupied by the terminating NULL.

Default

None

B.1.5.4 SCCOPT_XML_NULLREPLACECHAR

This option specifies a two-byte Unicode character that will be used to replace null characters if null path separators are being used. This option defaults to '/' and is valid for the SearchML 3.x, SearchHTML and SearchText output formats.

Handle Types

VTHDOC

Scope

Local

Data Type

VTWORD

Data

A two-byte Unicode character that will be used to replace null characters if null path separators are being used.

Default

0x002f = "/"

B.1.5.5 SCCOPT_XML_PAGEML_FLAGS

This option allows the developer to set flags that enable options unique to the PageML schema.

Handle Types

VTHDOC

Scope

Local

Data Type

VTDWORD

Data

One or more of the following values bitwise OR-ed together. Note that these flags are valid ONLY for the PageML output format:

- SCCEX_PAGEML_TEXTOUT: Include text in PageML's output.
- SCCEX_XML_NO_XML_DECLARATION: Exclude the XML declaration in PageML's output.

Default

- 0: All flags turned off.

B.1.5.6 SCCOPT_XML_PAGEML_PRINTERNAME

This option is Windows-specific. It is used to set which device context to use to render the pages.

It specifies, as a byte string, the name of the printer whose metrics should be used to calculate pagination information. If unspecified, the default printer will be used. The screen metrics of the system will be used if a printer is not specified and a default printer does not exist. As pagination is affected by the metrics of the device context and installed fonts, PageML XML output can vary between different systems and configurations.

Handle Types

VTHDOC

Scope

Local

Data Type

VTLPVOID

Data

A null-terminated single-byte string for the name of the printer which is the device context that should be used to render pages.

Default

- NULL
PageML uses the Windows default printer.

B.1.5.7 SCCOPT_XML_SEARCHML_CHAR_ATTRS

This option allows the developer to track character attributes contained in the input document and choose which are output to tags in the XML document produced.

Handle Types

VTHDOC

Scope

Local

Data Type

VTDWORD

Data

One or more of the following values bitwise OR-ed together. Note that not all flags are valid for all Search Export output formats.

- `SCCEX_XML_SEARCHML_ALLCAPS`: Valid for the SearchML 3.x output formats only.
- `SCCEX_XML_SEARCHML_BOLD`: Valid for the SearchML 3.x and SearchHTML output formats only.
- `SCCEX_XML_SEARCHML_DUNDERLINE`: Valid for the SearchML 3.x and SearchHTML output formats only.
- `SCCEX_XML_SEARCHML_HIDDEN`: Not valid for the PageML output format.
- `SCCEX_XML_SEARCHML_ITALIC`: Valid for the SearchML 3.x and SearchHTML output formats only.
- `SCCEX_XML_SEARCHML_OCE`: When this flag is set, an attribute named `oce` is added either to `<p>` or `<r>` elements as appropriate. (This flag does not affect `<unmapped>` elements, which will always have an `oce` attribute.) The value of the attribute is a hex representation of the character set. The value is defined by our core technology, `SO_ANSIUNKNOWN` for instance. Possible values for this attribute appear in the `vtchars.h` header file. Valid for the SearchML 3.x output formats only.
- `SCCEX_XML_SEARCHML_OUTLINE`: Valid for the SearchML 3.x output formats only.
- `SCCEX_XML_SEARCHML_REVISIONADD`: Valid for all output formats. When set, causes added text to be output and appropriately marked.
- `SCCEX_XML_SEARCHML_REVISIONDELETE`: Valid for all output formats. When set, causes deleted text to be output and appropriately marked.
- `SCCEX_XML_SEARCHML_SMALLCAPS`: Valid for the SearchML 3.x output formats only.
- `SCCEX_XML_SEARCHML_STRIKEOUT`: Valid for the SearchML 3.x output formats only.
- `SCCEX_XML_SEARCHML_UNDERLINE`: Valid for the SearchML 3.x and SearchHTML output formats only.

Default

- 0: All flags turned off.

B.1.5.8 `SCCOPT_XML_SEARCHML_FLAGS`

This option allows the developer to set flags that enable options unique to the following SearchML formats: SearchML 3.x, SearchHTML and SearchText.

This option is not valid for the PageML output format, although there is a similar PageML-specific option (`SCCOPT_XML_PAGEML_FLAGS`) that includes similar flags.

Handle Types

VTHDOC

Scope

Local

Data Type

VTDWORD

Data

One or more of the following values bitwise OR-ed together. Note that not all flags are valid for all Search Export output formats:

- **SCCEX_ANNOTATIONS**: When set, revised or annotated text will be designated as such. An "annotation" is a note or comment that goes along with a document, but is not really part of the document itself. Examples would be comments, footnotes, slidenotes, etc. Valid only for the SearchML 3.x output formats.
- **SCCEX_XML_ENABLEERRORINFO**: When this flag is set, SearchML will output an <error> element if an error occurs while processing the main document or any sub-documents. The <error> element has one required attribute, code, which will be a hex value of the error code. The contents of the element will be a string with the description of the error returned from DAGetErrorString. Valid only for the SearchML 3.1 and later output formats.
- **SCCEX_IND_GENERATED**: Includes data not originally stored as text in the input document. This can be important content the user would see when viewing the document in the original application (time and owner information in archives, numbers in spreadsheets/databases, etc.).
- **SCCEX_IND_GENERATESYSTEMMETADATA**: When this flag is set, system metadata will be generated. This text is "generated" and part of the document properties, so it will be affected by **SCCEX_IND_GENERATED** and **SCCEX_IND_SUPPRESSPROPERTIES**. This information is gathered through system calls and may adversely affect performance. Valid only for the SearchML 3.x output formats.
- **SCCEX_IND_SS_CELLINFO**: When this flag is set, SearchML will output a <cell> element that will encapsulate data from each non-empty cell in a spreadsheet. (NOTE: Numeric cells are considered empty unless **SCCEX_IND_GENERATED** is enabled.) The <cell> element will have a required attribute start which will give the location of the cell. It will also have an optional attribute end which will be used to indicate a merged cell. Both the start and end attributes will be in the form RowColumn where the Row will be a letter and Column will be a number (for example <cell start="A1">). Valid only for the SearchML 3.x output formats.
- **SCCEX_IND_SUPPRESSPROPERTIES**: Document properties are not produced. Not valid for the PageML output format.
- **SCCEX_METADATAONLY**: Produce only metadata.
- **SCCEX_PRODUCEURLS**: Produce URL and Book Mark information when it is available. Valid only for the SearchHTML and SearchML 3.x output formats.
- **SCCEX_XML_EMBEDDINGS**: Include embeddings.
- **SCCEX_XML_NO_XML_DECLARATION**: Exclude the XML declaration. Valid only for the SearchML 3.x output formats.
- **SCCEX_XML_PRODUCEOBJECTINFO**: When this flag is set, information for use with **IOTYPE_OBJECT** will be included in the <document> element. The information will correspond to the fields in the **SCCDAOBJECT** structure. Valid only for the SearchML 3.x output format.

- **SCCEX_XML_PSTYLENAMES:** Include paragraph style name references as an attribute of paragraph tags. Valid only for the SearchML 3.x output formats.
- **SCCEX_XML_SKIPSTYLES:** When possible, skip processing the style information. This should result in better performance, but certain output will no longer be available. When this flag is set and an appropriate filter is selected, character attributes, paragraph attributes, font names, and PDF Map Problem warnings will be unavailable, even if they have been requested.

Note: This will only work with optimized input filters, but Microsoft Office, PDF, RTF, MSG, Mime, and HTML are included in the optimized list.

- **SCCEX_XML_SUPPRESSARCHIVESUBDOCS:** Subdocuments in archives are not processed.
- **SCCEX_XML_SUPPRESSATTACHMENTS:** Attachments are not processed.

Default

- 0: All flags turned off.

B.1.5.9 SCCOPT_XML_SEARCHML_OFFSET

The value of this option is a Boolean that if set to `TRUE` will include offset information in the SearchML output according to the schema. If the option is set to `FALSE`, no offset information is produced.

Handle Types

VTHDOC, VTHEXPORT

Scope

Local

Data Type

VTBOOL

Default

FALSE

B.1.5.10 SCCOPT_XML_SEARCHML_PARA_ATTRS

This option allows the developer to track paragraph attributes contained in the input document and, optionally, include them in the XML output. All lengths are measured in twips. The values that appear in the SearchML output are the values that apply to the first content encountered in a given paragraph. For example, if the character height changes after the initial content in a paragraph, that change will be ignored. Left and first line indents are measured relative to the left page margin. The right indent is measured relative to the right page margin.

This option only affects SearchML output. The option is not valid for the SearchHTML, SearchText and PageML output flavors.

Handle Types

VTHDOC

Scope

Local

Data Type

VTDWORD

Data

One or more of the following values bitwise OR-ed together:

- SCCEX_XML_SEARCHML_SPACING
- SCCEX_XML_SEARCHML_HEIGHT
- SCCEX_XML_SEARCHML_LEFTINDENT
- SCCEX_XML_SEARCHML_RIGHTINDENT
- SCCEX_XML_SEARCHML_FIRSTINDENT

Default

- 0: All flags turned off.

B.1.5.11 SCCOPT_XML_SEARCHML_UNMAPPEDTEXT

This option allows for the production of unmapped text (the original code points from the input document). A new <unmapped> element will be produced to enclose this text. The <unmapped> element will contain base64-encoded text. It will also contain two attributes. "OCE" will contain a hex value representing the character set. "font" will contain a string value of the original font name. This is necessary for non-standard encodings such as wingdings or webdings. This option is only valid in the SearchML 3.2 (and higher) schema.

Handle Type

VTHDOC

Scope

Local

Data Type

VTDWORD

Data

One of the following values:

- SCCEX_XML_JUST_UNMAPPEDTEXT: Output just the unmapped text
- SCCEX_XML_NO_UNMAPPEDTEXT: Don't output any unmapped text.
- SCCEX_XML_BOTH_UNMAPPEDTEXT: Output both the original and the unmapped text.

Default

- SCCEX_XML_NO_UNMAPPEDTEXT

B.1.6 File System

This section discusses file system options.

B.1.6.1 SCCOPT_IO_BUFFERSIZE

This set of three options allows the user to adjust buffer sizes to tailor memory usage to the machine's ability. The numbers specified in these options are in kilobytes. These are advanced options that casual users of Search Export may ignore.

Handle Type

NULL, VTHDOC

Scope

Global

Data Type

SCCBUFFEROPTIONS Structure

Data

A buffer options structure

B.1.6.1.1 SCCBUFFEROPTIONS Structure

```
typedef struct SCCBUFFEROPTIONStag
{
    VTDWORD dwReadBufferSize;    /* size of the I/O Read buffer
                                in KB */
    VTDWORD dwMMapBufferSize;    /* maximum size for the I/O
                                Memory Map buffer in KB */
    VTDWORD dwTempBufferSize;    /* maximum size for the memory-
                                mapped temp files in KB */
    VTDWORD dwFlags;            /* use flags */
} SCCBUFFEROPTIONS, *PSCCBUFFEROPTIONS;
```

Parameters

- **dwReadBufferSize:** Used to define the number of bytes that will read from disk into memory at any given time. Once the buffer has data, further file reads will proceed within the buffer until the end of the buffer is reached, at which point the buffer will again be filled from the disk. This can lead to performance improvements in many file formats, regardless of the size of the document.
- **dwMMapBufferSize:** Used to define a maximum size that a document can be and use a memory-mapped I/O model. In this situation, the entire file is read from disk into memory and all further I/O is performed on the data in memory. This can lead to significantly improved performance, but note that either the entire file can be read into memory, or it cannot. If both of these buffers are set, then if the file is smaller than the dwMMapBufferSize, the entire file will be read into memory; if not, it will be read in blocks defined by the dwReadBufferSize.
- **dwTempBufferSize:** The maximum size that a temporary file can occupy in memory before being written to disk as a physical file. Storing temporary files in memory can boost performance on archives, files that have embedded objects or attachments. If set to 0, all temporary files will be written to disk.
- **dwFlags**

- SCCBUFOPT_SET_READBUFSIZE 1
- SCCBUFOPT_SET_MMAPBUFSIZE 2
- SCCBUFOPT_SET_TEMPBUFSIZE 4

To set any of the three buffer sizes, set the corresponding flag while calling `dwSetOption`.

Default

The default settings for these options are:

- `#define SCCBUFOPT_DEFAULT_READBUFSIZE 2`: A 2KB read buffer.
- `#define SCCBUFOPT_DEFAULT_MMAPBUFSIZE 8192`: An 8MB memory-map size.
- `#define SCCBUFOPT_DEFAULT_TEMPBUFSIZE 2048`: A 2MB temp-file limit.

Minimum and maximum sizes for each are:

- `SCCBUFOPT_MIN_READBUFSIZE 1`: Read one Kbyte at a time.
- `SCCBUFOPT_MIN_MMAPBUFSIZE 0`: Don't use memory-mapped input.
- `SCCBUFOPT_MIN_TEMPBUFSIZE 0`: Don't use memory temp files
- `SCCBUFOPT_MAX_READBUFSIZE 0x003fffff`, `SCCBUFOPT_MAX_MMAPBUFSIZE 0x003fffff`, `SCCBUFOPT_MAX_TEMPBUFSIZE 0x003fffff`: These maximums correspond to the largest file size possible under the 4GB DWORD limit.

B.1.6.2 SCCOPT_TEMPDIR

From time to time, the technology needs to create one or more temporary files. This option sets the directory to be used for those files.

It is recommended that this option be set as part of a system to clean up temporary files left behind in the event of abnormal program termination. By using this option with code to delete files older than a predefined time limit, the OEM can help to ensure that the number of temporary files does not grow without limit.

Note: This option will be ignored if `SCCOPT_REDIRECTTEMPFILE` is set.

Handle Types

NULL, VTHDOC

Scope

Global

Data Type

SCCUTTEMPDIRSPEC structure

B.1.6.2.1 SCCUTTEMPDIRSPEC Structure This structure is used in the `SCCOPT_TEMPDIR` option.

`SCCUTTEMPDIRSPEC` is a C data structure defined in `scvhw.h` as follows:

```
typedef struct SCCUTTEMPDIRSPEC
```

```

{
    VTDWORD    dwSize;
    VTDWORD    dwSpecType;
    VTBYTE     szTempDirName[SCCUT_FILENAMESIZE];
} SCCUTTEMPDIRSPEC, * LPSCCUTTEMPDIRSPEC;

```

There is currently a limitation. `dwSpecType` describes the contents of `szTempDirName`. Together, `dwSpecType` and `szTempDirName` describe the location of the source file. The only `dwSpecType` values supported at this time are:

- `IOTYPE_ANSIPATH`: Windows only. `szTempDirName` points to a NULL-terminated full path name using the ANSI character set and FAT 8.3 (Win16) or NTFS (Win32 and Win64) file name conventions.
- `IOTYPE_UNICODEPATH`: Windows only. `szTempDirName` points to a NULL-terminated full path name using the Unicode character set and NTFS file name conventions. Note that the length of the path name is limited to `SCCUT_FILENAMESIZE` bytes, or $(\text{SCCUT_FILENAMESIZE} / 2)$ double-byte Unicode characters.
- `IOTYPE_UNIXPATH`: X Windows on UNIX platforms only. `szTempDirName` points to a NULL-terminated full path name using the system default character set and UNIX path conventions.

Specifically not supported at this time is `IOTYPE_REDIRECT`.

Parameters

- `dwSize`: Set to `sizeof(SCCUTTEMPDIRSPEC)`.
- `dwSpecType`: `IOTYPE_ANSIPATH`, `IOTYPE_UNICODE` or `IOTYPE_UNIXPATH`
- `szTempDirName`: The path to the directory to use for the temporary files. Note that if all `SCCUT_FILENAMESIZE` bytes in the buffer are filled, there will not be space left for file names.

Default

The system default directory for temporary files. On UNIX systems, this is the value of environment variable `$TMP`. On Windows systems, it is the value of environment variable `%TMP%`.

B.1.6.3 SCCOPT_DOCUMENTMEMORYMODE

This option determines the maximum amount of memory that the chunker may use to store the document's data, from 4 MB to 1 GB. The more memory the chunker has available to it, the less often it needs to re-read data from the document.

Handle Types

NULL, VTHDOC

Scope

Global

Data Type

VTDWORD

Parameters

- `SCCDOCUMENTMEMORYMODE_SMALLEST 1 - 4MB`

- SCCDOCUMENTMEMORYMODE_SMALL 2 - 16MB
- SCCDOCUMENTMEMORYMODE_MEDIUM 3 - 64MB
- SCCDOCUMENTMEMORYMODE_LARGE 4 - 256MB
- SCCDOCUMENTMEMORYMODE_LARGEST 5 - 1 GB

Default

SCCDOCUMENTMEMORYMODE_SMALL 2 - 16MB

B.1.6.4 SCCOPT_REDIRECTTEMPFILE

This option is set when the developer wants to use redirected IO to completely take over responsibility for the low level IO calls of the temp file.

Handle Types

NULL, VTHDOC

Scope

Global (not persistent)

Data Type

VTLPVOID: pCallbackFunc

Function pointer of the redirect IO callback.

Redirect call back function:

```
typedef
{
    VTDWORD (* REDIRECTTEMPFILECALLBACKPROC)
    (HIOFILE *phFile,
    VTVOID *pSpec,
    VTDWORD dwFileFlags);
```

There is another option to handle the temp directory, SCCOPT_TEMPDIR. Only one of these two can be set by the developer. The SCCOPT_TEMPDIR option will be ignored if SCCOPT_REDIRECTTEMPFILE is set. These files may be safely deleted when the Close function is called.

B.2 Search Export SOAP Options

These options are available to the developer when using the export engine through the Transformation Server API.

This chapter details the Web Services implementation of options in Transformation Server. However, there are links to API-specific information for the C and JAVA client interfaces to the technology within each of the following sections.

B.2.1 How Options Work

An option is defined by an identifier and an associated value. The identifier (hOptions) indicates what particular option is being specified. The option value data must be in a form that conforms to the set of supported data types.

Note that it is not necessarily an error to specify options that are not understood by the export engine, but some transformation engines may require that certain options be specified.

B.2.2 Character Mapping

This section discusses character mapping options.

B.2.2.1 defaultInputCharset

This option is used in cases where Oracle Outside In cannot determine the character set used to encode the text of an input file. When all other means of determining the file's character set are exhausted, Oracle Outside In will assume that an input document is encoded in the character set specified by this option. This is most often used when reading plain-text files, but may also be used when reading HTML or PDF files.

When the "extended test for text" is enabled (see [Section B.2.4.2, "extendedTestForText"](#)), this option will still apply to plain-text input files that are not identified as EBCDIC or Unicode.

This option supersedes the fallbackFormat option for selecting the character set assumed for plain-text files. For backwards compatibility, use of deprecated character-set -related values is still currently supported for fallbackFormat, though internally such values will be translated into equivalent values for the defaultInputCharset. As a result, if an application were to set both options, the last such value set for either option will be the value that takes effect.

Data Type

DefaultInputCharSet

Data

The SOAP representation of the character set to use, from the values in defaultInputCharSetEnum.

B.2.2.2 unmappableCharacter

This option selects the character used when a character is not a valid Unicode character, or does not conform to the XML specification for valid characters. This option takes the Unicode value for the replacement character. If you are using the PageML output format, this option is only valid if the textOutOn option is set.

Data Type

xsd:unsignedShort

Data

The Unicode value for the character to use.

Default

- 0xffffd

Links

- C Client Implementation: XSD_unsignedShort
- JAVA Client Implementation: UnsignedShort

B.2.3 Output

This section discusses output options.

B.2.3.1 preferOITRendering

This option is only valid on the Linux (Red Hat and Suse) and Solaris Sparc platforms.

This option is only valid when PageML is the output format.

When this option is set to true, the technology will attempt to use its internal graphics code to render fonts and graphics. When set to false, the technology will render images using the operating system's native graphics subsystem (X11 on UNIX/Linux platforms). Note that this option only works when at least one of the appropriate output solutions is present. For example, if the UNIX \$DISPLAY variable does not point to a valid X Server, but the OSGD and/or WV_GD modules required for the Oracle Outside In output solution exist, Oracle Outside In will default to the Oracle Outside In rendering code. The option will fail if neither of these output solutions is present.

It is important for the system to be able to locate useable fonts when this option is set to true. Only TrueType fonts (*.ttf or *.ttc files) are currently supported. To ensure that the system can find them, make sure that the environment variable GDFONTPATH includes one or more paths to these files. If the variable GDFONTPATH can't be found, the current directory is used. If fonts are called for and cannot be found, Search Export will exit with an error. Also note that when copying Windows fonts to a UNIX system, the font extension for the files (*.ttf or *.ttc) must be lowercase, or they will not be detected during the search for available fonts. Oracle does not provide fonts with any Oracle Outside In product.

If preferOITRendering is set in a particular instance of tsagent, it cannot be changed in that agent until the agent is terminated.

Data Type

xsd:boolean

Data

One of the following values:

- true: Use the technology's internal graphics rendering code to produce bitmap output files whenever possible.
- false: Use the operating system's native graphics subsystem.

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.4 Input Handling

This section discusses input handling options.

B.2.4.1 fallbackFormat

This option controls how files are handled when their specific application type cannot be determined. This normally affects all plain-text files, because plain-text files are generally identified by process of elimination, for example, when a file isn't identified as having been created by a known application, it is treated as a plain-text file.

A number of values that were formerly allowed for this option have been deprecated. Specifically, the values that selected specific plain-text character sets are no longer to be used. Instead, applications should use the [defaultInputCharset](#) option for such functionality.

Data Type

FallbackFormatEnum

Data

One of the following values:

- `fallbackToText`: Unidentified file types will be treated as text files.
- `noFallbackFormat`: Oracle Outside In will not attempt to process files whose type cannot be identified. This will include text files. When this option is selected, an attempt to process a file of unidentified type will cause Oracle Outside In to return an error value of `SCCERR_UNSUPPORTEDFORMAT`.

Default

- `ASCII-8`

Links

- C Client Implementation: `OIT_FallbackFormatEnum`
- JAVA Client Implementation: `FallbackFormatEnum`

B.2.4.2 extendedTestForText

This option affects how an input file's internal format (application type) is identified when the file is first opened by the Oracle Outside In technology. When the extended test flag is in effect, and an input file is identified as being either 7-bit ASCII, EBCDIC, or Unicode, the file's contents will be interpreted as such by the export process.

The extended test is optional because it requires extra processing and cannot guarantee complete accuracy (which would require the inspection of every single byte in a file to eliminate false positives.)

Data Type

xsd:boolean

Data

One of the following values:

- `false`: This is the default value. When this is set, standard file identification behavior occurs.
- `true`: If set, the File Identification code will run an extended test on all files that are not identified.

Default

- false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.4.3 ignorePassword

This option can disable the password verification of files where the contents can be processed without validation of the password. If this option is not set, the filter should prompt for a password if it handles password-protected files.

As of Release 8.4.0, only the PST and MDB Filters support this option.

Data Type

xsd:boolean

Data

- true: Ignore validation of the password
- false: Prompt for the password

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.4.4 oleEmbeddings

Microsoft Powerpoint versions from 1997 through 2003 had the capability to embed OLE documents in the Powerpoint files. This option controls which embeddings are to be processed as native (OLE) documents and which are processed using the alternate graphic.

Note: The Microsoft Powerpoint application sometimes does embed known Microsoft OLE embeddings (such as Visio, Project) as an "Unknown" type. To process these embeddings, the processAll option is required. Post Office-2003 products such as Office 2007 embeddings also fall into this category.

Data Type

OleEmbeddingsEnum

Data

- processAll: Process all embeddings in the file.
- processNone: Process none of the embeddings in the file
- processStandard: Process embeddings that are known standard embeddings.

Default

processStandard

Links

- C Client Implementation: OIT_OleEmbeddingsEnum
- JAVA Client Implementation: OleEmbeddingsEnum

B.2.4.5 parseXMPMetaData

Adobe's Extensible Metadata Platform (XMP) is a labeling technology that allows you to embed data about a file, known as metadata, into the file itself. This option enables parsing of the XMP data into normal OIT document properties. Enabling this option may cause the loss of some regular data in premium graphics filters (such as Postscript), but won't affect most formats (such as PDF).

Data Type

xsd:boolean

Data

- true: This setting enables parsing XMP.
- false: This setting disables parsing XMP.

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.4.6 reorderBIDI

This option controls whether or not the PDF filter will attempt to reorder bidirectional text runs so that the output is in standard logical order as used by the Unicode 2.0 and later specification. This additional processing will result in slower filter performance according to the amount of bidirectional data in the file.

Data Type

xsd:boolean

Data

- true: The PDF filter uses standard ordering.
- false: The PDF filter will attempt to reorder bidirectional text runs.

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.4.7 timezone

This option allows the user to define an offset to GMT that will be applied during date formatting, allowing date values to be displayed in a selectable time zone. This option affects the formatting of numbers that have been defined as date values (e.g., most dates in spreadsheet cells). This option will not affect dates that are stored as text.

Note: This option does not apply for spreadsheet files.

Data Type

xsd:int

Data

Integer parameter from -96 to 96, representing 15-minute offsets from GMT. To query the operating system for the time zone set on the machine, specify the numeric value of 61440 (0xF000 in hexadecimal).

Default

- 0: GMT time

Links

- C Client Implementation: XSD_int
- JAVA Client Implementation: Integer

B.2.4.8 extractXMPMetaData

Adobe's Extensible Metadata Platform (XMP) is a labeling technology that allows you to embed data about a file, known as metadata, into the file itself. This option enables the XMP feature, which does not interpret the XMP metadata, but passes it straight through without any interpretation.

Data Type

xsd:boolean

Data

- true
- false

Default

- false

B.2.4.9 htmlCondCommentIE5On

This option allows you to display content customized for Internet Explorer 5.

Data Type

xsd_boolean

Default

0: off

Links

C Client Implementation: VTBOOL

JAVA Client Implementation: boolean

B.2.4.10 htmlCondCommentIE6On

This option allows you to display content customized for Internet Explorer 6.

Data Type

xsd_boolean

Default

0: off

Links

C Client Implementation: VTBOOL

JAVA Client Implementation: boolean

B.2.4.11 htmlCondCommentIE7On

This option allows you to display content customized for Internet Explorer 7.

Data Type

xsd_boolean

Default

0: off

Links

C Client Implementation: VTBOOL

JAVA Client Implementation: boolean

B.2.4.12 htmlCondCommentIE8On

This option allows you to display content customized for Internet Explorer 8.

Data Type

xsd_boolean

Default

0: off

Links

C Client Implementation: VTBOOL

JAVA Client Implementation: boolean

B.2.4.13 htmlCondCommentIE9On

This option allows you to display content customized for Internet Explorer 9.

Data Type

xsd_boolean

Default

0: off

Links

C Client Implementation: VTBOOL

JAVA Client Implementation: boolean

B.2.4.14 htmlCondCommentAllOn

This option allows you to display all conditional comments.

Data Type

xsd:boolean

Default

0: off

Links

C Client Implementation: VTBOOL

JAVA Client Implementation: boolean

B.2.5 Compression

This section discusses compression options.

B.2.5.1 allowLZW

This option can disable access to any files using Lempel-Ziv-Welch (LZW) compression, such as .GIF files, .ZIP files or self-extracting archive (.EXE) files containing "shrunk" files. Attempts to read such files when this option is enabled will fail and return the error SCCERR_UNSUPPORTEDCOMPRESSION.

The following is a list of file types affected when this option is disabled:

- GIF files
- TIF files using LZW compression
- PDF files that use internal LZW compression
- TAZ and TAR archives containing files that are identified as FI_UNIXCOMP
- ZIP and self-extracting archive (.EXE) files containing "shrunk" files
- Postscript files using LZW compression

Although this option can disable access to files in ZIP or EXE archives stored using LZW compression, any files in such archives that were stored using any other form of compression will still be accessible.

Data Type

xsd:boolean

Data

- true: LZW compressed files will be read and written normally.
- false: LZW compressed files will not be read or written.

Default

true

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6 XML

This section pertains to XML options.

B.2.6.1 allCapsOn

When set, causes capitalized text to be output and appropriately marked. Valid for the SearchML 3.x output formats only.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.2 boldOn

When set, causes bold text to be output and appropriately marked. Not valid for the SearchText and PageML output formats.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.3 cellInfoOn

When set, SearchML will output a <cell> element that will encapsulate data from each non-empty cell in a spreadsheet. (Note: Numeric cells are considered empty unless FI DOCS NO HP BUILDING(3.7) is enabled.) The <cell> element will have a required attribute start which will give the location of the cell. It will also have an optional attribute end which will be used to indicate a merged cell. Both the start and end attributes will be in the form RowColumn where the Row will be a letter and Column will be a number (for example, <cell start="A1">). Valid only for the SearchML 3.x output formats.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.4 changeNumberToTextOn

Includes data not originally stored as text in the input document. This can be important content the user would see when viewing the document in the original application (time and owner information in archives, numbers in spreadsheets/databases, etc.). Valid for all output formats.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.5 documentPropertiesOn

When set, document properties are included in the output. Default value is false. Not valid for the PageML output format.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.6 doubleUnderlineOn

When set, causes double-underlined text to be included in the output and appropriately marked. Not valid for the SearchText and PageML output formats.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.7 embeddingsOn

Include embeddings. Not valid for the PageML output format.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.8 errorInfoOn

When this flag is set, SearchML will output an <error> element if an error occurs while processing the main document or any sub-documents. The <error> element has one required attribute, code, which will be a hex value of the error code. The contents of the element will be a string with the description of the error returned from DAGetErrorString. Valid only for the SearchML 3.x output formats.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.9 generateSystemMetaDataOn

When this flag is set, system metadata will be generated. This text is "generated" and is part of the document properties. This information is gathered through system calls and may adversely affect performance. Valid only for the SearchML 3.x output formats.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.10 hiddenOn

Include hidden text in the output. Not valid for the PageML output format.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.11 italicOn

Include italic text in the output. Not valid for the SearchText and PageML output formats.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.12 metadataOnlyOn

Produce only metadata. Not valid for the PageML output formats.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.13 originalCharsetOn

When this option is set, an attribute named `oce` is added either to `<p>` or `<r>` elements as appropriate. The value of the attribute is a hex representation of the character set.

The value is defined by our core technology, `SO_ANSIUNKNOWN` for instance. Possible values for this attribute appear in the `vtchars.h` header file. Valid for the SearchML 3.x output formats only.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.14 outlineOn

Include outlined text in the output. Valid for the SearchML 3.x output formats only.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.15 produceURLsOn

Produce URL information when it is available. Valid for the SearchML 3.x and SearchHTML output formats only.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.16 revisionAddOn

When set, causes added text to be output and appropriately marked. Valid for all output formats.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.17 revisionDeleteOn

When set, causes deleted text to be output and appropriately marked. Valid for all output formats.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.18 revisionsOn

When set, revised or annotated text will be designated as such. Valid only for the SearchML 3.x output format.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.19 smallCapsOn

When set, causes text in small caps to be output and appropriately marked. Valid for the SearchML output format only.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean

- JAVA Client Implementation: Boolean

B.2.6.20 strikeoutOn

When set, causes strikethrough text to be output and appropriately marked. Valid for the SearchML 3.x output formats only.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.21 underlineOn

When set, causes underlined text to be output and appropriately marked. Valid for the SearchML 3.x output formats only.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.22 xmlDefinitionMethod

This option determines whether Search Export will reference a SearchML or PageML schema, DTD, or no reference when generating output. This option is not valid when SearchText or SearchHTML is the output format.

Data Type

XMLDefinitionMethodEnum

Data

One of the following values:

- dtd: Document Type Definition (DTD)
- xsd: Extensible Schema Definition
- noDefinition: No XML definition reference

Default

noDefinition

Links

- C Client Implementation: OIT_XmlDefinitionMethodEnum
- JAVA Client Implementation: XmlDefinitionMethodEnum

B.2.6.23 xmlDefinitionLocation

This option allows the developer to set a particular file as the XML definition reference.

If the [xmlDefinitionMethod](#) option is set to `xsd` or `dtd`, the value of this option will be used to reference the schema or DTD, respectively.

Data Type

xsd:string

Data

A UTF-8 encoded string specifying the location of an `xsd` or `dtd` file. If using the C API, this string must be a null-terminated array of single-byte characters.

Default

None

Links

- C Client Implementation: XSD_string
- JAVA Client Implementation: String

B.2.6.24 nullReplacementCharacter

This option specifies a two-byte Unicode character that will be used to replace null characters if null path separators are being used. This option defaults to `'/'` and is valid for the SearchML 3.x, SearchHTML and SearchText output formats.

Data Type

xsd:unsignedShort

Data

A two-byte Unicode character that will be used to replace null characters if null path separators are being used.

Default

0x002f = `"/`

Links

- C Client Implementation: XSD_unsignedShort
- JAVA Client Implementation: UnsignedShort

B.2.6.25 printerName

This option is Windows-specific. It is used to set which device context to use to render the pages.

It specifies, as a byte string, the name of the printer whose metrics should be used to calculate pagination information. If unspecified, the default printer will be used. The screen metrics of the system will be used if a printer is not specified and a default printer does not exist. As pagination is affected by the metrics of the device context and installed fonts, PageML XML output can vary between different systems and configurations.

Data Type

xsd:string

Data

A null-terminated single-byte string for the name of the printer which is the device context that should be used to render pages.

Default

- null: PageML uses the Windows default printer.

Links

- C Client Implementation: XSD_string
- JAVA Client Implementation: String

B.2.6.26 paragraphStyleNamesOn

Include paragraph style name references as an attribute of paragraph tags. Valid for the SearchML 3.x output formats only.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.27 includeTextOffsets

The value of this option is a Boolean that if set to true will include offset information in the SearchML output according to the schema. If the option is set to false, no offset information is produced.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean

- JAVA Client Implementation: Boolean

B.2.6.28 paragraphAttributes

This option allows the developer to track paragraph attributes contained in the input document and, optionally, include them in the XML output. All lengths are measured in twips. The values that appear in the SearchML output are the values that apply to the first content encountered in a given paragraph. For example, if the character height changes after the initial content in a paragraph, that change will be ignored. Left and first line indents are measured relative to the left page margin. The right indent is measured relative to the right page margin.

Data Type

ParagraphAttributes

Data

The paragraphAttributes option is a complexType data structure composed of Boolean variables, which may be switched on or off in any combination. The variables are:

- spacing
- height
- leftIndent
- rightIndent
- firstIndent

Default

- 0: All flags set to false.

Links

- C Client Implementation: OIT_ParagraphAttributes
- JAVA Client Implementation: ParagraphAttributes

B.2.6.29 unmappedText

This option allows for the production of unmapped text (the original code points from the input document). A new <unmapped> element will be produced to enclose this text. The <unmapped> element will contain base64-encoded text. It will also contain two attributes. "OCE" will contain a hex value representing the character set. "Font" will contain a string value of the original font name. This is necessary for non-standard encodings such as wingdings or webdings. This option is only valid in the SearchML 3.2 (and higher) schema.

Data Type

SearchMLUnmappedTextEnum

Data

One of the following values:

- justUnmappedText: Output just the unmapped text
- noUnmappedText: Don't output any unmapped text.
- bothUnmappedText: Output both the original and the unmapped text.

Default

- noUnmappedText

Links

- C Client Implementation: OIT_SearchMLUnmappedTextEnum
- JAVA Client Implementation: SearchMLUnmappedTextEnum

B.2.6.30 suppressArchiveSubDocsOn

Subdocuments in archives are not processed. Not valid for the PageML output format.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.31 suppressAttachmentsOn

Attachments are not processed. Not valid for the PageML output format.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.32 textOutOn

This option is valid only for the PageML output format.

When set to true, include text in the PageML output.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.6.33 xmlDeclarationOff

Exclude the XML declaration. Not valid for the SearchText and SearchHTML output formats.

Data Type

xsd:boolean

Default

false

Links

- C Client Implementation: XSD_boolean
- JAVA Client Implementation: Boolean

B.2.7 File System

This section applies to file system options.

B.2.7.1 fileAccess

This option supplies information to OIT when information is required to open an input file. This information may be the password of the file or a support file location.

Further information about how Transformation Server implements this option will be forthcoming.

B.2.7.2 readBufferSize

Used to define the number of bytes that that will read from disk into memory at any given time. Once the buffer has data, further file reads will proceed within the buffer until the end of the buffer is reached, at which point the buffer will again be filled from the disk. This can lead to performance improvements in many file formats, regardless of the size of the document.

Data Type

xsd:unsignedInt

Data

The size of the buffer in kilobytes.

Default

2

Links

- C Client Implementation: XSD_unsignedInt
- JAVA Client Implementation: UnsignedInt

B.2.7.3 memoryMappedInputSize

Used to define a maximum size that a document can be and use a memory-mapped I/O model. In this situation, the entire file is read from disk into memory and all further I/O is performed on the data in memory. This can lead to significantly improved performance, but note that either the entire file can be read into memory, or

it cannot. If both of these buffers are set, then if the file is smaller than the `dwMMapBufferSize`, the entire file will be read into memory, if not, it will be read in blocks defined by the `dwReadBufferSize`.

Data Type

xsd:unsignedInt

Data

The size of the buffer in kilobytes.

Default

8192

Links

- C Client Implementation: XSD_unsignedInt
- JAVA Client Implementation: UnsignedInt

B.2.7.4 tempBufferSize

The maximum size that a temporary file can occupy in memory before being written to disk as a physical file. Storing temporary files in memory can boost performance on archives, files that have embedded objects or attachments. If set to 0, all temporary files will be written to disk.

Data Type

xsd:unsignedInt

Data

The size of the buffer in kilobytes.

Default

2048

Links

- C Client Implementation: XSD_unsignedInt
- JAVA Client Implementation: UnsignedInt

Symbols

\$DISPLAY, 3-7
\$HOME, 3-9
\$LD_LIBRARY_PATH, 3-9
\$LIBPATH, 3-9
\$ORIGIN, 3-8
\$SHLIB_PATH, 3-9

A

allCapsOn, B-31
allowLZW, B-30
Architectural Overview, 1-3

B

batch_process_sx, 8-2
boldOn, B-31

C

C/C++ Options, B-1
cellInfoOn, B-31
CGI programs, 3-7
changeNumberToTextOn, B-32
Character Mapping, B-1, B-23
Compression, B-11, B-30
Copyright, 1-6

D

DACloseDocument, 4-5
DACloseTreeRecord, 4-16
DADeInit, 4-3
DAGetErrorString, 4-9
DAGetFileId, 4-7
DAGetFileIdEx, 4-8
DAGetObjectInfo, 4-9
DAGetOption, 4-7
DAGetTreeCount, 4-10
DAGetTreeRecord, 4-10
DAInitEx, 4-2
DAOpenDocument, 4-3
DAOpenRandomTreeRecord, 4-12
DAOpenTreeRecord, 4-12
DARetrieveDocHandle, 4-6

DASaveInputObject, 4-13
DASaveRandomTreeRecord, 4-15
DASaveTreeRecord, 4-14
DASetFileAccessCallback, 4-17
DASetOption, 4-6
DASetStatCallback, 4-16
Data Access Common Functions, 4-1
DATREENODELOCATOR, 4-13, 4-16
defaultInputCharset, B-23
Deprecated Functions, 4-2
Directory Structure, 1-5
documentPropertiesOn, B-32
doubleUnderlineOn, B-32

E

embeddingsOn, B-33
environment variables, 3-9
 \$DISPLAY, 3-7
 \$HOME, 3-9
 \$LD_LIBRARY_PATH, 3-9
 \$LIBPATH, 3-9
 \$SHLIB_PATH, 3-9
errorInfoOn, B-33
EXCloseExport, 5-2
EXExportStatus, 5-3
export, 8-2
 Main Window, 8-3
Export Functions, 5-1
EXRunExport, 5-3
exsimple, 8-4
extendedTestForText, B-25
extract_archive, 8-4
extractXMPMetaData, B-28

F

fallbackFormat, B-25
File System, B-19, B-42
fileAccess, B-42

G

generateSystemMetaDataOn, B-33
GLIBC and Compiler Versions, 3-12
graphic types, 3-7

H

hiddenOn, B-34
How to Use Search Export, 1-5
HP-UX Compiling and Linking, 3-10
HP-UX on Itanium (64 bit), 3-10
HP-UX on RISC, 3-10
HP-UX on RISC (64 bit), 3-10
htmlCondCommentAllOn, B-30
htmlCondCommentIE5On, B-28
htmlCondCommentIE6On, B-29
htmlCondCommentIE7On, B-29
htmlCondCommentIE8On, B-29
htmlCondCommentIE9On, B-29

I

IBM AIX (32-bit pSeries), 3-11
IBM AIX Compiling and Linking, 3-10
IBM AIX PPC (64-bit), 3-11
ignorePassword, B-26
Implementation Issues, 7-1
includeTextOffsets, B-39
Input Handling, B-3, B-24
Introduction, 1-1
IOClose, 6-2
IOGENSECONDARY and IOGENSECONDARYW
Structures, 6-8
IOGetInfo, 6-5
IOGETINFO_GENSECONDARY, 6-9
IORead, 6-3
IOSeek, 6-4
IOSPEARCHIVEOBJECT Structure, 4-5
IOSPECLINKEDOBJECT Structure, 4-4
IOTell, 6-5
IOWrite, 6-3
italicOn, B-34

L

Licensing, A-1
Linux 32-bit, including Linux PPC, 3-15
Linux 64-bit, 3-15
Linux Compiling and Linking, 3-11
Linux zSeries, 3-15

M

memoryMappedInputSize, B-42
metadataOnlyOn, B-34
Motif Libraries, 3-11

N

NSF Support, 2-2, 3-2
nullReplacementCharacter, B-38

O

OLE2, 3-8
oleEmbeddings, B-26

Oracle Solaris (SPARC) 64, 3-16
Oracle Solaris Compiling and Linking, 3-15
Oracle Solaris SPARC, 3-15
Oracle Solaris x86, 3-16
originalCharsetOn, B-34
outlineOn, B-35
Output, B-2, B-24

P

PageML, 1-2
paragraphAttributes, B-40
paragraphStyleNamesOn, B-39
parseXMPMetaData, B-27
preferOITRendering, B-24
printerName, B-38
process_isolate, 8-4
produceURLsOn, B-35

Q

query folders, 3-7

R

readBufferSize, B-42
reorderBIDI, B-27
revisionAddOn, B-35
revisionDeleteOn, B-36
revisionsOn, B-36
Running in 24x7 Environments, 7-1
Running in Multiple Threads or Processes, 7-1
Runtime Search Path, 3-8

S

Sample Applications, 8-1
SCCBUFFEROPTIONS Structure, B-19
SCCCA_TREENODELOCATOR, 4-13, 4-16
SCCDAOBJECT Structure, 4-5
SCCDATREENODE Structure, 4-11
SCCOPT_ARCFULLPATH, B-10
SCCOPT_DEFAULTINPUTCHARSET, B-1
SCCOPT_DOCUMENTMEMORYMODE, B-21
SCCOPT_ENABLEALLSUBOBJECTS, B-11
SCCOPT_EXTRACTXMPMETADATA, B-3
SCCOPT_FALLBACKFORMAT, B-4
SCCOPT_FIFLAGS, B-5
SCCOPT_FILTERLZW, B-11
SCCOPT_FORMATFLAGS, B-5
SCCOPT_HTML_COND_COMMENT_MODE, B-9
SCCOPT_IGNORE_PASSWORD, B-6
SCCOPT_IO_BUFFER_SIZE, B-19
SCCOPT_LOTUSNOTESDIRECTORY, B-6
SCCOPT_PARSEXMPMETADATA, B-7
SCCOPT_PDF_FILTER_DROPHYPHENS, B-10
SCCOPT_PDF_FILTER_REORDER_BIDI, B-7
SCCOPT_PROCESS_OLE_EMBEDDINGS, B-8
SCCOPT_REDIRECTTEMPFILE, B-22
SCCOPT_RENDERING_PREFER_OIT, B-2
SCCOPT_SYSTEMFLAGS, B-6

SCCOPT_TEMPDIR, B-20
SCCOPT_TIMEZONE, B-9
SCCOPT_UNMAPPABLECHAR, B-2
SCCOPT_XML_DEF_METHOD, B-12
SCCOPT_XML_DEF_REFERENCE, B-12
SCCOPT_XML_NULLREPLACECHAR, B-13
SCCOPT_XML_PAGEML_FLAGS, B-13
SCCOPT_XML_PAGEML_PRINTERNAME, B-14
SCCOPT_XML_SEARCHML_CHAR_ATTRS, B-14
SCCOPT_XML_SEARCHML_FLAGS, B-15
SCCOPT_XML_SEARCHML_OFFSET, B-17
SCCOPT_XML_SEARCHML_PARA_ATTRS, B-17
SCCOPT_XML_SEARCHML_
UNMAPPEDTEXT, B-18
SCCUTTEMPDIRSPEC Structure, B-20
Search Export Options, B-1
SearchHTML, 1-3
SearchML, 1-2
SearchText, 1-3
smallCapsOn, B-36
SOAP Options, B-22
Status Callback Function, 4-16
strikeoutOn, B-37
suppressArchiveSubDocsOn, B-41
suppressAttachmentsOn, B-41
xsxample, 8-2

T

tempBufferSize, B-43
textOutOn, B-41
timezone, B-28
Tree Node Locator, 4-13, 4-16

U

underlineOn, B-37
UNIX
API Libraries, 3-2
Changing Resources, 3-9
Character Sets, 3-7
Engine Libraries, 3-4
Environment Variables, 3-9
Filter and Export Filter Libraries, 3-4
Information Storage, 3-6
Installation, 3-1
Libraries and Structure, 3-2
OLE2, 3-8
Premier Graphics Filters, 3-5
Runtime Considerations, 3-7
Signal Handling, 3-8
Support Libraries, 3-3
Unix
X server, 3-7
UNIX Implementation Details, 3-1
unmappableCharacter, B-23
unmappedText, B-40
Using Redirected IO, 6-1

V

vector graphics, 3-7

W

What's New in Release 8.4.1, 1-1
Windows
API DLLs, 2-2
Changing Resources, 2-9
Character Sets, 2-7
Engine Libraries, 2-4
Filter and Export Filter Libraries, 2-4
Installation, 2-1
Libraries and Structure, 2-2
Options and Information Storage, 2-6
Premier Graphics Filters, 2-5
Support DLLs, 2-3
Windows Implementation Details, 2-1

X

XML, B-11, B-31
xmlDeclarationOff, B-42
xmlDefinitionLocation, B-38
xmlDefinitionMethod, B-37
xxredir, 8-4

Z

z/OS Compiling and Linking, 3-16

