Oracle® Secure Enterprise Search

Administrator's Guide

11*g* Release 2 (11.2.2.2)

E35070-15

May 2016

ORACLE®

Oracle Secure Enterprise Search Administrator's Guide, 11*g* Release 2 (11.2.2.2)

E35070-15

Primary Author: Rajesh Bhatiya

Contributor: Arun Kalyaanasundaram, Aswini Krishnan, Ching-tai Chiu, Cindy Hsin, Diego Iglesias, Greg Brunet, Hiroshi Koide, Luke Wang, Marvin Huang, Michael Kuo, Mohammad Faisal, Mohammed Qureshi, Pranav Agarwal, Pranav Singh, Ravi Palakodety, Rishishankar Rengasamy, Sachin Bhatkar, Sana Karam, Santhosh Thomas, Shashi Anand, Shijun Cheng, Stefan Buchta, Steve Yang, Susanta Mohapatra, Takeshi Yabe, Thomas Chang, Tommy Mo, Valarie Moore, Wei We Chen, Yan Zhao, Yaohua Ma

Contents

3 Understanding the Oracle SES Search User Interface

4 Understanding Crawling

Part II Creating Data Sources

6 Configuring Access to Built-in Sources

7 Configuring Access to Content Management Sources

9   Configuring Access to Applications Sources

Part III   Advanced Topics

10   Customizing the Search User Interface

11 Security in Oracle Secure Enterprise Search

12   Administering Oracle SES Instances

13 Oracle Secure Enterprise Search APIs

Preface

The *Oracle Secure Enterprise Search Administrator's Guide* explains how to administer Oracle Secure Enterprise Search instances. You will learn how to set up a variety of information sources, crawl and index those sources, and customize the search results.

This Preface contains these topics:

- Audience
- Documentation Accessibility
- Related Documents
- Conventions

Audience

Oracle Secure Enterprise Search Administrator's Guide is intended for administrators and application developers who perform the following tasks:

- Install and configure Oracle Secure Enterprise Search
- Administer Oracle Secure Enterprise Search
- Develop Oracle Secure Enterprise Search 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 that have purchased support 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 about Oracle Secure Enterprise Search, refer to the following resources:

- *Oracle Secure Enterprise Search Release Notes*

  Provides version information and identifies known issues.

- *Oracle Secure Enterprise Search Installation Guides*

  Discuss installation requirements and tips, and provides information on how to get started using Oracle Secure Enterprise Search.

- *Oracle Secure Enterprise Search Administration API Guide*

  Provides an introduction to the various interfaces to the Administration API.

- *Oracle Secure Enterprise Search Administration Java API Reference*

  Describes the classes and methods in the Oracle SES Administration Java API for efficiently managing multiple instances of Oracle SES.

- *Oracle Secure Enterprise Search Java API Reference*

  Describes the classes and methods in the Oracle SES Java API.

Up-to-date Release Notes are posted on Oracle Technology Network (OTN). You must register online before using OTN. Registration is free and can be done at this location:

http://www.oracle.com/technetwork/community/join/overview/index.htm

If you have a user name and password for OTN, then you can go directly to the documentation section of OTN at this location:

http://www.oracle.com/technetwork/indexes/documentation/index.html

Conventions

The following text conventions are used in this document:

Convention	Meaning
/	A slash separates levels of a directory path. On Windows, use a backslash (\) instead of a slash (/).
boldface	Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.
italic	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.
oracle_base	*oracle_base* is the root of the directory structure where all the Oracle Database products are installed.
oracle_home	*oracle_home* refers to the directory where the database component specific to Oracle SES software is installed. The default *oracle_home* directory is *oracle_base*/product/11.2.0/sesdb_1.
mw_home	*mw_home* refers to the directory where the middle tier component of Oracle SES software (that is, Oracle WebLogic Server, Oracle SES server, Oracle Fusion Middleware common files) is installed.
ses_home	*ses_home* refers to the directory where the Oracle SES server is installed. When Oracle SES 11.2.2.2 software is installed along with the WebLogic Server middle tier, the *ses_home* directory is *mw_home*/Oracle_SES1.
wls_home	*wls_home* refers to the directory where the WebLogic Server components are installed. When Oracle SES 11.2.2.2 software is installed along with the WebLogic Server, the *wls_home* directory is *mw_home*/wlserver.

Convention	Meaning
wls_domain_home	*wls_domain_home* refers to the directory where multiple instances of a WebLogic Server middle tier are created. The default *wls_domain_home* directory is *mw_home*/user_projects/domains.
ses_domain_name	*ses_domain_name* refers to the directory where the WebLogic Server instance specific to Oracle SES is created under *wls_domain_home* directory. The default *ses_domain_name* directory is search_domain.
java_home	*java_home* refers to the directory where JDK is installed in the Oracle SES middle tier. When Oracle SES 11.2.2.2 software is installed along with the WebLogic Server middle tier, the *java_home* directory is *ses_home*/jdk.

Note: In the previous Oracle SES releases, the Oracle SES home was referred to as *ORACLE_HOME*. In this release, the Oracle SES home is referred to as *ses_home*. The *ses_home* directory is created under the *mw_home* directory. When Oracle SES 11.2.2.2 is installed along with the WebLogic Server middle tier, the *ses_home* directory is *mw_home*/Oracle_SES1.

In this release *oracle_home* refers to the directory where the database component specific to Oracle SES software is installed.

What's New

This chapter provides the overview of the new and changed features in Oracle Secure Enterprise Search (Oracle SES) 11*g*, Release 2 (11.2.2.2), and provides pointers to additional information.

New and Changed Features for Release 11.2.2.2

Oracle SES 11.2.2.2 includes the following new and changed features:

Architectural Changes

- Oracle SES 11.2.2.2 supports open architecture, that is, an Oracle SES application can have multiple middle tier components across multiple systems. Therefore, you must ensure that the custom plug-in jar files are accessible across all the Oracle SES instance middle tiers.

  In the earlier Oracle SES release, a custom plug-in jar file was required to be stored in the *ses_home*/search/lib/plugins directory, but this is not required in the current Oracle SES release. You can now store a custom plug-in jar file in any directory, but you must refer it using its absolute file path in the Oracle SES application. See "Extensible Crawler Plug-in Framework" on page 1-7 for more information.

Changed Features

- You can now use the searchctl command only when both the Oracle Database and the WebLogic Server middle tier are installed as part of the Oracle SES software installation.

- The Oracle SES administrative user name is SEARCHSYS in this release. For the earlier Oracle SES release (11.1.2.2), the administrative user name is EQSYS.

New Features

- Faceted Navigation: You can now use faceted navigation in query application to refine your search results based upon various categories of data. See "Configuring and Using Facets in Search Results" on page 5-13 for more information.

- Hard Sort: You can now order the search results based on the sortable search attributes. See "Configuring Sort Criteria using Sortable Attributes (Absolute Sort)" on page 5-22 for more information.

- Sitemap Crawl: You can now use Sitemap crawl for crawling Web sources efficiently. See "Sitemap Crawl" on page 4-7 for more information.

- Auto Suggest: You can now get the query suggestions as you type query terms in the search box in the query application. See "Auto Suggestions" on page 3-3 for more information.

- Personalization: You can now configure user specific preference profiles for controlling the behavior of query results in the query application. See "Personalized Search using Preference Profiles" on page 3-16 for more information.

- Push Crawl: You can now use the Push crawl mechanism for crawling and indexing documents in almost real time, that is, as soon as the documents are available in a source system. See "Understanding the Push Type Crawling Process" on page 4-30 for more information.

- Tagging: You can now use tags to classify documents into multiple categories, thus increasing the relevancy of search results. See "Configuring and Using Tags in Search Results" on page 5-17 for more information.

- Changes to the default MIME types processing for data sources: The configuration for default MIME types that are processed for each data source type has changed in this release. See "Document Types" on page 4-5 for more information.

- SharePoint 2010 support: Oracle SES now has a connector for SharePoint 2010 sources. See "Setting Up Microsoft SharePoint Sources" on page 7-11 for more information.

- Crawling of OAM-protected Web data sources support: SES now supports crawling of OAM-protected Web sites using HTTP basic authentication.

- The wildcard search is not enabled by default in Oracle SES. You can enable the wildcard search in Oracle SES by configuring it using either the Administration GUI or the Administration API. See "Basic Search" on page 3-1 for more information.

- You can change the master encryption key with the `searchctl rollover_key` command. See "Changing the Master Encryption Key" on page 11-52.

Desupported Features

- Backup and recovery using Administration GUI: You now cannot use the Administration GUI for the Oracle SES backup and recovery operations. See "Performing Backup and Recovery" on page 12-1 for more information.

- Disk space management: You now cannot specify the crawler log directory and the cache directory using Administration GUI and Administration API. The pre-seeded storage area entries for these directories are removed in this release. The Administration API objects for disk space management - `alert`, `spaceCalculator`, and `task` - are also not supported in this release.

- Deploying Oracle SES instance as a portlet: You now cannot deploy Oracle SES instance as a portlet.

- Changing the Oracle SES administrator password using Administration GUI: In this release, Oracle SES administrator is a WebLogic Server identity store user, and hence the administrator password can be changed only by using the standard interface provided by the WebLogic Server. You now cannot change the administrator password using the Oracle SES Administration GUI. See "Changing the Administrator Password" on page 11-2 for more information.

- Configuration files `crawler.dat`, `search.properties`, `ranking.xml`, `drivers.properties`, and `QueryPlan.xml`: These configuration files are not supported in this release, and you can now specify these configurations by using either the Administration GUI or the Administration API or both.

Configuration file desupported in this release	Description	Corresponding Administration GUI page in this release	Corresponding Administration API objects in this release
crawler.dat	Crawler configuration	Global Settings - Crawler Configuration	crawlerSettings
search.properties	Query application configuration	Global Settings - Query UI Configuration	queryUIConfig
ranking.xml	Query ranking configurations	None	relevanceRanking
drivers.properties	Database connector configuration	Global Settings - Source Types - Database	source object of Database type
QueryPlan.xml	Query application configuration	Global Settings - Query UI Configuration	queryUiConfig

■ The following source types are not supported in this release:

  – FileNet Content Engine

  – FileNet Image Services

  – Hummingbird

  – IBM DB2

  – Open Text Livelink

  – EMC Documentum eRoom

Part I

Learning the Basics

This part provides the information you need to administer Oracle SES instances. It contains the following chapters:

1

Introduction to Oracle Secure Enterprise Search

This chapter describes the basic components of Oracle Secure Enterprise Search: the sources, crawler, and user interfaces. It contains the following topics:

- Special-Use Licensing
- Overview of Oracle Secure Enterprise Search
- Source Types
- Oracle Secure Enterprise Search Components
- Oracle Secure Enterprise Search Features
- Oracle Secure Enterprise Search Language Model

Special-Use Licensing

Oracle Secure Enterprise Search (Oracle SES) is a complete stacked application. Oracle Database 11*g* Release 2 (11.2.0.3) Enterprise Edition (EE) is installed with Oracle SES. Use of Oracle Database EE is restricted to storing and managing the search index, metadata, cache, and Oracle SES configuration information. Oracle WebLogic Server 10*g* (10.3.6.0) is included with Oracle SES. This embedded version is provided solely to run the Oracle SES user interfaces and APIs.

The Oracle SES home software use is restricted to support the Oracle SES database repository, and no other databases created using the Oracle SES executables are supported. Oracle SES connectors listed on the Oracle price list may be licensed separately to use with the Oracle SES installation.

Some connectors shipped with Oracle SES require additional licensing fees. Contact Oracle sales for details.

Overview of Oracle Secure Enterprise Search

Oracle Secure Enterprise Search enables a secure, high quality, easy-to-use search across all enterprise information assets. Key features include:

- The ability to search and locate public, private and shared content across Intranet web-servers, databases, files on local disks or file servers, IMAP e-mail, document management systems, applications, and portals
- Highly secure crawling, indexing, and searching
- A simple, intuitive search interface leading to an excellent user experience

- Excellent search quality, with the most relevant items for a query shown first, even when the query spans diverse public and private data sources

- Analytics on search results and usage patterns

- Sub-second query performance

- Ease of administration and maintenance, leveraging existing IT expertise

    See Also:

    - *Oracle Secure Enterprise Search Installation Guide* for requirements, tips, and information on getting started using Oracle SES

    - Oracle Technology Network (OTN) for updated information on the known issues, code samples, and best practices:

        http://www.oracle.com/technetwork/search/oses/overview/index.html

    - The *Oracle Secure Enterprise Search Release Notes* has version information and known issues

Source Types

A collection of information is called a source. Each source has a type that identifies where the information is stored, such as on a Web site or in a database table. Oracle SES provides several built-in source types and an architecture for adding new, custom types.

Additionally, Oracle SES provides access to more third-party data repositories than any other enterprise search engine, without requiring you to generate any additional coding. While these data sources are classified as user-defined source types, they are available the same as the built-in source types. This guide organizes these user-defined source types into content management sources, collaboration sources, and applications sources.

Oracle SES also provides authorization cache sources for facilitating access to secure data.

Built-in Sources

- **Web**: Represents the content on a specific Web site. Web sources facilitate maintenance crawling of specific Web sites.

- **Table**: Represents content in a table or view in Oracle Database.

- **File**: The set of documents that can be accessed through the file system protocol.

- **E-mail**: Derives content from e-mails sent to a specific e-mail address. When Oracle SES crawls an e-mail source, it collects e-mail from all folders set up in the e-mail account, including Drafts, Sent Items, and Trash e-mails.

- **Mailing list**: Derives its content from e-mails sent to a specific mailing list.

- **OracleAS Portal**: Lets you search across multiple OracleAS Portal repositories, such as Web pages, files on disk, and pages on other OracleAS Portal instances.

- **Federated**: Enable you to share content across multiple Oracle SES instances.

Content Management Sources

- EMC Documentum Content Server

- Microsoft SharePoint

- Oracle Content Database

- Oracle Content Server (formerly Stellent Content Server)

You may need to install client libraries and obtain a license from the vendor for some content sources to work. For example, EMC Documentum requires installation of a compatible version of Documentum Foundation Classes (DFC), which is a Java library, on the computer running Oracle SES. Oracle SES does not ship with DFC.

Collaboration Sources

- IBM Lotus Notes

- Microsoft Exchange

- Microsoft NT File Systems (NTFS)

- Oracle Calendar

- Oracle Collaboration Suite E-Mail

Oracle Applications Sources

- Database

- Oracle E-Business Suite

- Siebel

- Oracle Fusion

- Oracle WebCenter

Authorization Sources

- User Authorization Cache

- Federated User Authorization Cache

> **See Also:** *Oracle Secure Enterprise Search Release Notes* for a list of supported platforms

Oracle Secure Enterprise Search Components

Oracle SES includes the following components:

- Oracle Secure Enterprise Search Administration GUI

- Oracle Secure Enterprise Search Crawler

- Oracle Secure Enterprise Search APIs

Oracle Secure Enterprise Search Administration GUI

The Oracle Secure Enterprise Search Administration GUI enables you to manage and monitor Oracle SES components using a browser-based interface. These are among the tasks that you perform:

- Define sources and crawling scope

- Configure the search application

- Monitor crawl progress and search quality

- Customize search results

See Also:

- "Understanding the Oracle SES Administration GUI" on page 2-2

- *Oracle SES Administration Tutorial* at
  `http://www.oracle.com/pls/topic/lookup?ctx=E35215-01&id=tut_ses_administration` for help understanding common administrator tasks

- Oracle SES Administration GUI Help

Oracle Secure Enterprise Search Crawler

Oracle SES uses a crawler to collect data from the sources. The Oracle SES crawler is a Java process activated by a schedule. When activated, the crawler spawns a configurable number of processor threads that fetch information from various sources and index the documents. This **index** is used for searching **sources**.

The crawler maps links and analyzes relationships. Whenever the crawler encounters embedded non-HTML, or non-textual documents during the crawling, it automatically detects the document type, and filters and indexes the document.

Figure 1–1 shows the crawler in relation to other Oracle SES components and a variety of data sources.

Figure 1–1   Crawler Collecting Information for Oracle SES

Oracle Secure Enterprise Search APIs

Oracle Secure Enterprise Search provides several APIs. For example, with the Web Services API, you can integrate Oracle SES search capabilities into your search application. You can also customize the default Oracle SES ranking to create a more relevant search result list for your enterprise or configure clustering for customized applications.

The Crawler Plug-in API enables you to create a custom secure crawler plug-in (or **connector**) to meet your requirements. The Document Service API accepts input from documents and performs some operation on it. For example, you could create a document service for auditing or to show custom metatags.

See Also:

- Chapter 13, "Oracle Secure Enterprise Search APIs"
- *Oracle Secure Enterprise Search Administration API Guide*
- *Oracle Secure Enterprise Search Java API Reference*

Oracle Secure Enterprise Search Features

Information in an enterprise can be spread across Web pages, databases, mail servers or other collaboration software, document repositories, file servers, and desktops. Oracle SES searches all your data through the same interface. Oracle SES is fully globalized and works with many languages including Chinese, Japanese, Korean, Arabic, and Hebrew.

This section introduces a few of the features in Oracle SES. It includes the following topics:

- Secure Search
- Federated Search
- Extensible Crawler Plug-in Framework

> **See Also:** Chapter 4, "Understanding Crawling" for more features relating to the crawler

Secure Search

Much of the information within an organization is publicly accessible. Anyone is allowed to view it. Therefore, it is relatively easy for a **crawler** to find and index that information.

However, there are other sources that are protected. These protected sources might be viewable only by certain users or groups of users. For example, while users can search in their own e-mail folders, they should not be able to search anyone else's e-mail.

For protected sources, the Oracle SES crawler indexes documents with the proper access control list. When end users perform a search, only documents that they have privileges to view are returned.

> **See Also:** "Enabling Secure Search" on page 11-13

Federated Search

Oracle SES can search multiple Oracle SES instances with their own document repositories and indexes. It provides a unified framework to search the different repositories that are crawled, indexed, and maintained separately.

Federated search allows a single query to run across all Oracle SES instances. It aggregates the search results to show one unified result list to the user. User credentials are passed along with the query so that each federation endpoint can authenticate the user against its own document repository.

Figure 1–2 illustrates the federation architecture and two options for an end user to connect through a browser to Oracle SES. Option 1 allows users to connect their browsers directly to Oracle SES using the end-user graphical interface. Option 2 retrieves results from Oracle SES through Web Services after arbitrary post-processing, such as changing the look-and-feel or embedding the results in a page. For this option, the browser connects to remote applications, which connect to the Web Services API.

Figure 1–2   Federation Architecture
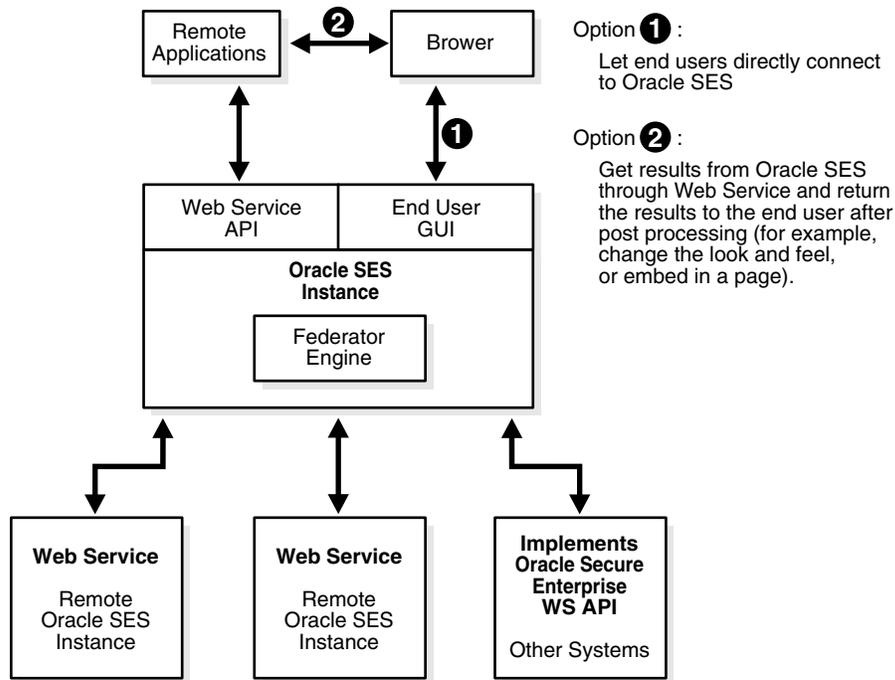
Extensible Crawler Plug-in Framework

Oracle SES provides an extensible crawler plug-in framework that lets you crawl and index proprietary document repositories. The Crawler Plug-in API enables you to create a custom secure crawler plug-in to meet your requirements. You can also create a custom identity plug-in and a custom authorization plug-in for crawling a data source. You can also update or delete a custom plug-in.

Oracle SES 11.2.2.2 supports open architecture, that is, an Oracle SES application can have multiple middle tier components across multiple systems. You must ensure that the custom plug-in jar files must be accessible across all the Oracle SES instance middle tiers.

Thus, if all the middle tiers have a shared file system, then the custom plug-in jar files must be placed at a location which is accessible to all the middle tiers. If all the middle tiers do not have a shared file system, then you must create the same directory structure across all the middle tiers and copy the custom plug-in jar files to all these locations. For example, if the path specified for a custom plug-in jar file is `/app/install/plugins/custom_plugin.jar`, then the same directory structure must be created for all the middle tiers.

> **Note:**   In the earlier Oracle SES release (Oracle SES 11.1.2.2), a custom plug-in jar file was required to be stored in the `ses_home/search/lib/plugins` directory, but this is not required in Oracle SES 11.2.2.2 release. You can now store a custom plug-in jar file in any directory, but you must refer it using its absolute file path in the Oracle SES application.

The following are the various custom plug-ins that need to be accessible across all the middle tiers for an Oracle SES instance:

- Source Type plug-in

- Authorization plug-in

- Identity plug-in

- Document Service plug-in

- URL Rewriter plug-in

For example:

- Document service for "Default summarizer Doc service" uses a stop words directory. The default stop words directory is *ses_home*/search/lib/plugins/doc/extractor/phrasestopwords.

- Document Service for "ImageDocumentService" uses an attributes configuration file. Its default directory location is *ses_home*/search/lib/plugins/doc/ordim/config/attr-config.xml.

- Database source type uses an XML query file.

- Federated User Authorization Cache source type uses an XML remote cache configuration file.

> **Note:** Documentum Content Server source type uses the following property files:
>
> - search/lib/plugins/dcs/dfc.properties
>
> - search/lib/plugins/dcs/dcsothers/dfc.properties
>
> As these files are used by Documentum APIs, they must be stored in the directory *ses_home*/search/lib/plugins.

> **See Also:**
>
> - "Oracle Secure Enterprise Search Java SDK" on page 13-39
>
> - The Oracle Secure Enterprise Search home page at http://www.oracle.com/technetwork/search/oses/overview/index.html for updated information on known issues, code samples, and best practices

Oracle Secure Enterprise Search Language Model

Different components of Oracle SES, such as crawler, Query application, Administration GUI, Administration API, Query API, and ODL support different set of languages.

This section provides information about the languages supported by these components. This section includes the following topics:

- Languages Supported by Crawler

- Languages Supported by the Query Application and Query API

- Languages Supported by the Administration GUI and Administration API

- Languages Supported by ODL Logs

Languages Supported by Crawler

An Oracle SES data source can contain documents in different languages, for example, one document can be in English, while other can be in Japanese. Oracle SES associates a single document with a single language.

A data source can explicitly specify a language for a document. Oracle SES recognizes only the document languages that are specified using the ISO 639-1 standard, but not all of the languages.

The Administration GUI can also be used to set one of the following document languages as a default language for a data source - Arabic, Chinese, Czech, Danish, Dutch, English, Finnish, French, German, Greek, Hebrew, Hungarian, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Romanian, Russian, Slovak, Spanish, Swedish, and Turkish.

If a default language is not set for a data source using Administration GUI, then the crawler can automatically detect a document's language by reading its metadata. The crawler can detect the following document languages automatically - Arabic, Chinese, Danish, Dutch, English, French, German, Greek, Japanese, Korean, Italian, Lao, Portuguese, Spanish, Thai, and Tibetan.

> **Note:** The automatic language detection feature of Oracle SES is enabled by default for all the data source types, except the file data source type. You can also disable the automatic language detection feature using Administration GUI. See "Language Detection" on page 4-9 for more information.

The crawler determines a document's language by performing the following checks in the mentioned order:

- From the HTTP response header *Content-Language* for a Web data source

- From the HTML *Language* meta tag, for example, *<meta name="Language" content ="en">*

- From the HTML *content-language* meta tag, for example, *<meta http-equiv="content-language" content="fr">*

- From the *LANGUAGE* column of a Table data source

- From the language specified for a crawler Plug-in

- Using the automatic language detection feature (if this feature is enabled)

- From the default language specified for a data source

Languages Supported by the Query Application and Query API

The default query application supports the following languages - Arabic, Catalan, Chinese, Czech, Danish, German, Greek, English, Spanish, Finnish, French, Hungarian, Italian, Japanese, Korean, Dutch, Norwegian, Polish, Portuguese, Romanian, Russian, Slovak, Swedish, Thai, and Turkish.

The default query application obtains the client-side language, territory, and character-set information from the Web browser. The language information is then passed as is to the Query API, only if CJK characters (Chinese, Japanese, and Korean) are not present in the query, otherwise, the language is set to the most likely one among the CJK in the Query API.

A query in Oracle SES can be specified using only one language. A document that is not in the same language as specified in the query, is either not searched at all or is assigned lower relevancy score in the query search result depending upon the Oracle SES configuration.

> **Note:** The language specific behavior of the default query application may change in the future releases of Oracle SES without notice.

Oracle SES Query API supports different languages for different types of search as described in Table 1–1.

Table 1–1   Languages Supported by Query API for Different Types of Search

Search Type	Languages Supported
Stemming search	English, Dutch, French, German, Italian, Spanish
Fuzzy search	English, Dutch, French, German, Italian, Korean, Spanish, Chinese, and Japanese. For Chinese and Japanese languages, only the VGRAM search is supported. The VGRAM search is insensitive to word boundaries.
Composite word search	German
Alternate spelling search	German

Languages Supported by the Administration GUI and Administration API

The Administration GUI supports the following languages - Chinese, English, French, German, Italian, Japanese, Korean, Portuguese, and Spanish.

You can translate the search attribute names to the following languages using Administration GUI and Administration API - English, Arabic, Brazilian Portuguese, Catalan, Czech, Danish, Dutch, Finnish, French, German, Greek, Hebrew, Hungarian, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Romanian, Russian, Simplified Chinese, Slovak, Spanish, Swedish, Thai, Tradition Chinese, and Turkish.

Languages Supported by ODL Logs

Oracle SES uses ODL for logging the warning and error messages into log files. Oracle SES supports the following languages for logging these messages - English, German, Spanish, Franch, Italian, Japanese, Korean, Brazilian Portuguese, Simplified Chinese, and Traditional Chinese.

2

Getting Started with the Oracle SES Administration GUI

This chapter provides a brief introduction to using Oracle Secure Enterprise Search (Oracle SES). More information about the Oracle SES Administration GUI is provided later in this book and in the online Help.

This chapter contains the following topics:

- Getting Started Basics for the Oracle SES Administration GUI
- Understanding the Oracle SES Administration GUI
- Starting and Stopping Oracle SES Instance

Getting Started Basics for the Oracle SES Administration GUI

After you have successfully installed Oracle SES, you can define your data sources and start crawling them. The following are general procedures.

To open the Oracle SES Administration GUI:

1. Open a browser and enter the URL for the Oracle SES Administration GUI. This URL is provided after the installation and has the form
   `http://host:port/search/admin/index.jsp`

2. Log in with the user name `searchsys` and the password specified during installation.

To administer Oracle SES:

1. Define one or more sources for the data you want to search.

   For example, if your data is stored in Web pages, then select Web source.

2. Check the crawler progress and status on the Home - Schedules page. Click **Refresh Status**. From the status page, you can view the statistics of the crawl.

3. Monitor the search statistics on the Home - General page and the Home - Statistics page.

The following procedures expand these steps.

> **Note:** For specific information about a source, see Part II, "Creating Data Sources".

To create a source:

1. On the home page, select the **Sources** secondary tab to display the Sources page.

2. Select a source type.

3. Click **Create** to display the Create Source page.

4. Complete the page. Click **Help** for a description of each item.

5. Click **Create** or **Create & Customize**.

   A crawl schedule is automatically created along with the source. If you select **Start Crawling Immediately**, then the crawler starts crawling when you click **Create**.

To customize a source:

1. When creating a source, click **Create & Customize** on the Create Source page to display the Customize Source page.

   or

   After creating a source, click the **Edit** icon for the source on the Home - Sources page.

2. Click the subtabs and make the desired changes.

3. Click **Apply**.

To crawl and index an existing source:

1. On the home page, select the **Schedules** secondary tab to display the Crawler Schedules page.

2. Select the schedule.

3. To set up the frequency and other crawling options for one or more sources, click **Edit**. (Optional)

4. Click **Start** to start crawling.

5. To test whether users can search a source, click the **Search** link in the upper right of any page.

   This link displays the search page in a new window. The URL for **Search** has the form `http://host:port/search/query/search`.

Understanding the Oracle SES Administration GUI

The Oracle SES Administration GUI provides many options for managing and customizing Oracle SES to suit your enterprise. This section describes some tasks you can accomplish using the Oracle SES Administration GUI.

Home Tab

The **Home** tab consists of the **General**, **Sources**, **Schedules**, and **Statistics** secondary tabs.
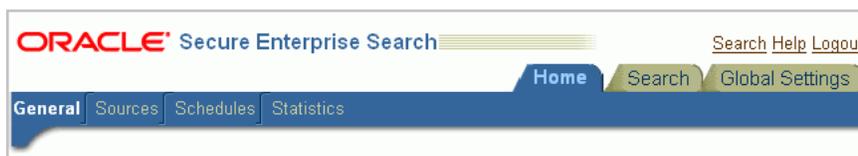
- **Home - General**

This is the home page for Oracle SES. The Summary section shows an overview of the system's search performance, both quality and speed, over the past seven days. The Failed Schedules section lists all schedules that have failed. A failed schedule is one in which the crawler encountered an irrecoverable error, such as an indexing error or a source-specific login error, and cannot proceed. A failed schedule could be because of a partial collection and indexing of documents.

- **Home - Sources**

  A collection of information is called a source. Each source has a type, such as a Web site or a database table. User-defined source types are created on the Global Settings - Source Types page. The list includes any available user-defined source types. You can create as many sources as you want.

- **Home - Schedules**

  This page lets you view, edit, create, delete, stop, or start a schedule. Schedules define the frequency at which the index is updated with information about each source.

- **Home - Statistics**

  This page provides numerous search and crawler statistics, such as most popular queries and crawler progress.

  Some statistics constantly show up-to-date information, while others are cached hourly to improve performance. The Last Refreshed time shows the actual time of the statistics displayed. Check the online help for each statistics page to confirm if the statistics are up-to-date or cached hourly.

Search Tab

The **Search** tab consists of the **Relevancy**, **Suggested Links**, **Suggested Content, Alternate Words**, and **Source Groups** secondary tabs. These pages help you improve search quality.

- **Search - Relevancy**

  Make important documents easier to find with relevancy boosting. Oracle SES lets you influence the order of documents in the result list for a particular search. For example, your company Web site could have a home page for documentation that should appear high in the results of any search for `documentation`.

- **Search - Suggested Links**

  Direct users to a particular Web site for a search string. For example, when users search for "Oracle SES documentation" or "Enterprise Search documentation" or "Search documentation", you could suggest `http://www.oracle.com/technology`. In the default search page, suggested links are displayed at the top of the search result list. This is especially useful to provide links to important Web pages that are not crawled by Oracle SES.

- **Search - Suggested Content**

  Suggest actual content (as opposed to links) to be displayed along with the result list. For example, when an end user searches for contact information on a

coworker, Oracle SES fetches the content from the suggested content provider and returns the contact information (e-mail address, phone number, and so on) for that person with the result list. Suggested content results appear in tabbed panes above the query results.

■ **Search - Alternate Words**

Use alternate words to suggest alternative search queries to users. This is useful for fixing common errors that users make in searching (for example, entering Oracel instead of Oracle). Also, synonyms can provide more relevant results; for example, cellular phones for cell phones or wireless phones. Additional uses for alternate keywords are for product code names and abbreviations.

■ **Search - Source Groups**

Source groups are groups of sources that can be searched together. A source group consists of one or more sources, and a source can be assigned to multiple source groups. On the basic Search page, users can browse source groups that the administrator created. If no source group is selected, then all documents are searched.

Global Settings Tab

The **Global Settings** tab includes links to configure settings for your Oracle SES environment.

This section describes some global configuration pages.

■ **Crawler Configuration**

This page configures global crawler settings, such as crawling depth, language, and maximum document size.

After a source has been created, you can define crawling parameters, such as URL boundary rules and crawling depth, for that source by editing that source on the Home - Sources page.

■ **Query Configuration**

This page includes several options, including the following: maximum number of results returned, display URL, spell checking, statistics collection, URL submission, federated search, and secure search.

- **Identity Management Setup**

  This page lets you set up connections between Oracle Secure Enterprise Search and any identity management system to validate and authenticate users. This is necessary for secure searches. Oracle SES uses an **identity plug-in** as an interface to an identity management system.

- **Configure Search Results List**

  This page lets you customize the look and feel of the search result list using XSLT and CSS style sheets.

- **Configure Clustering in Search Results**

  Real-time clustering organizes search results into various groups to provide end users with different views on the top results. Clustered documents within one group (called a cluster node) share the same common topics or property values. A cluster node with a large document set can be categorized into child cluster nodes, and a hierarchy cluster result tree is built. Users can navigate directly to a specific cluster node or refine their query by combining the original query and cluster results.

- **Configure Facets in Search Results**

  The facets are used in refining the query results in the query application by using the facet panel and the facet navigation bar. The faceted search provides the mechanism for filtering data in a search result based on the categorization and sub-categorization of data. The various categories of data can be considered as facets, and for each category, the sub-categories are called facet nodes. Thus, a hierarchy of facet tree is built. Users can navigate directly to a specific facet node or refine their query by combining the original query with the facet search results.

  The pages - Facets, Translate Facet Name, and Configure Facet Name - let you configure facets by providing the facet related operations, such as, creating, updating, deleting, enabling and disabling, setting visibility parameters, and translating facet names to different languages.

- **Configure Source Groups**

  This page lets you configure the display properties of facets, cluster trees, top-n sortable attributes, top-n groupable attributes, and Absolute sorting criteria (default sort conditions and sortable attributes) for a source group that is enabled to be displayed in the query application.

- **Configure Single Sign-On**

  This page configures one of the following Single Sign-On (SSO) authentication mechanisms for Oracle SES:

  - Oracle Access Manager (OAM)
  - Oracle Single Sign-On (OSSO)
  - Windows Native Authentication (WNA)

See Also:

- *Oracle SES Administration Tutorial* at http://www.oracle.com/pls/topic/lookup?ctx=E35215-01&id=t ut_ses_administration for help understanding common administrator tasks

- Oracle SES Administration GUI context sensitive online Help

- Oracle SES home page for updated information on known issues, code samples, and best practices:

  http://www.oracle.com/technetwork/search/oses/overview/in dex.html

Starting and Stopping Oracle SES Instance

You can start and stop the Oracle SES instance (Oracle SES server, Oracle ESS server, WebLogic server, and Database server) using any of the following methods:

- Using WebLogic Administration Console

- Using WebLogic Server Scripts

- Using `searchctl` Command

> **Note:** In the previous Oracle SES releases, the Oracle SES home was referred to as *ORACLE_HOME*. In this release, the Oracle SES home is referred to as *ses_home*. The *ses_home* directory is created under the *mw_home* directory. When Oracle SES 11.2.2.2 is installed along with the WebLogic Server middle tier, the *ses_home* directory is *mw_ home*/Oracle_SES1.
>
> In this release *oracle_home* refers to the directory where the database component specific to Oracle SES software is installed.
>
> Refer to "Conventions" on page xviii for more information.

Using WebLogic Administration Console

You can start and stop the Oracle SES middle tier components (WebLogic Server, Oracle ESS Server, and Oracle SES Server) using the WebLogic Administration Console.

Using WebLogic Server Scripts

You can use the following WebLogic Server scripts to start and stop the Oracle SES middle tier components (WebLogic Server, Oracle ESS Server, and Oracle SES Server):

Description	WebLogic Server Script
Start WebLogic Server	`wls_domain_home/ses_domain_name/startWebLogic.sh`
Stop WebLogic Server	`wls_domain_home/ses_domain_name/bin/stopWeblogic.sh`
Start Oracle SES server	`wls_domain_home/ses_domain_ name/bin/startManagedWebLogic.sh ses_servar_name`
Stop Oracle SES server	`wls_domain_home/ses_domain_ name/bin/stopManagedWebLogic.sh ses_server_name admin_ URL user_name password`

Description	WebLogic Server Script
Start Oracle ESS server	`wls_domain_home/ses_domain_name/bin/startManagedWebLogic.sh ess_server_name`
Stop Oracle ESS server	`wls_domain_home/ses_domain_name/bin/stopManagedWebLogic.sh ess_server_name admin_URL user_name password`

Using `searchctl` Command

You can use the `searchctl` command for starting and stopping the Oracle SES instance, only when both the Database and the WebLogic Server middle tier are installed as part of the Oracle SES software installation.

The `searchctl` executable is located in the *ses_home*/bin directory:

`$ses_home/bin/searchctl`

The `searchctl` command syntax is as follows.

Command	Description
`searchctl start`	Start Oracle SES server
`searchctl stop`	Shut down Oracle SES server
`searchctl restart`	Restart Oracle SES server
`searchctl start_midtier`	Start WebLogic Server middle tier
`searchctl stop_midtier`	Shut down WebLogic Server middle tier
`searchctl restart_midtier`	Restart WebLogic Server middle tier
`searchctl start_backend`	Start Database server
`searchctl stop_backend`	Shut down Database server
`searchctl restart_backend`	Restart Database server
`searchctl startall`	Start Oracle SES server, WebLogic Server middle tier, and Database server
`searchctl stopall`	Stop Oracle SES server, WebLogic Server middle tier, and Database server
`searchctl restartall`	Restart Oracle SES server, WebLogic Server middle tier, and Database server

On Linux and UNIX systems, you must provide the Oracle SES administrator password when running the `searchctl` command. If the middle tier port is less than 1024, then the `searchctl` command for starting or stopping the WebLogic Server middle tier, must be run by the root user.

> **Note:** The `searchctl stopall` command does not close active background tasks before initiating an immediate shutdown. This may raise internal errors, which you can ignore.

3

Understanding the Oracle SES Search User Interface

This chapter describes the syntax that you can use to search for documents using Oracle SES. It introduces the search interface and explains how you can alter the search results.

This chapter contains the following topics:

- Searching using Oracle Secure Enterprise Search
- Basic Search
- Submit URL
- Configuring the Display Options on the Search Page
- Attribute Filters
- Advanced Search
- Thesaurus-Based Search
- Alternate Words
- Restricted Search and the Browse Popup
- Personalized Search using Preference Profiles

> **See Also:** "Tuning Search Performance" on page 12-13

Searching using Oracle Secure Enterprise Search

You can submit a search request using Oracle SES in the following places:

- The search box of the Oracle SES Query Application
- The Search page of the Oracle SES Administration GUI
- A query string of the Oracle SES Web Services API

> **See Also:** Oracle SES Query Application Online Help

Basic Search

Figure 3–1 shows the default Oracle SES query application. You enter your search string and click the **Search** button, in the same way as for any other search engine.

Figure 3–1   Oracle SES Query Application

A search string can consist of one or more words. The search is case-insensitive. The search using wildcard characters (* and ?) is not enabled by default in Oracle SES. You can enable wildcard search in Oracle SES either by selecting the **Enable Wildcard Search** parameter on the Administration GUI page **Global settings - Query Configuration** or by using the `queryConfig` object of the Administration API. Refer to Table 3–1 for the syntax for using wildcard search in the query application.

> **Note:**   Enabling the wildcard search reduces the query performance in Oracle SES.

Clicking the **Search** button returns all matches for that search string. The search string match can be found in the content of the document or in an attribute value of the document, such as the author.

Figure 3–2 shows sample results for the search string `Secure Enterprise Search`. You can reorder the way results are presented using the **Group by** and **Sort by** lists.

Figure 3–2   Sample Search Results
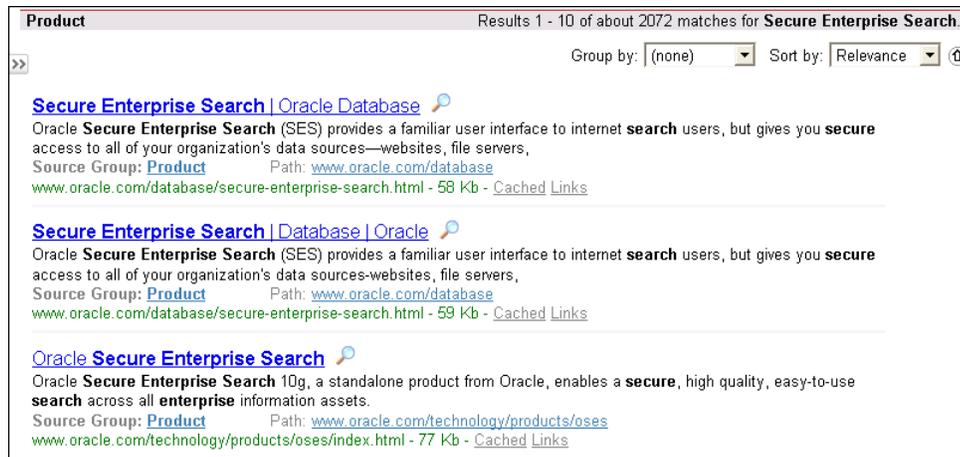
Keywords in Context

Oracle incorporates KWIC (keyword in context) as part of the search result. This has a size restriction of 4K. That is, if the searched keyword appears in the first 4K of a document, then the KWIC is shown for the search result. If the keyword appears after the first 4k, then no KWIC is shown.

Stemming Expansion

Oracle SES applies stemming to the query term. Stemming expands the term to other terms that share the same root. For example, [banks] returns documents containing the word banks, banking, or bank. Oracle SES uses stemming based on the language of query, which is determined by the language of browser in the default query application, or it is input by the caller in the query API. Implicit stemming expansion applies on individual terms in term search, in proximity search, and in attribute shortcut search for STRING attributes. Implicit stemming expansion does not apply to phrase search, and it can be turned off by enclosing the term in double quotes.

Alternate Words Expansion

Oracle SES can provide a suggestion for a query in the form of a "do you mean..." message for alternate word pairs. When you select the Auto-Expand option in the Oracle SES Administration GUI, Oracle SES automatically includes the alternate keyword as part of the search. An alternate keyword definition composed of multiple words is handled like a phrase in the search. For example, if the alternate word pair is "RAC" and "Real Application Clusters," then a query for "RAC" returns documents containing the word "RAC" or the phrase "Real Application Clusters."

If the search administrator has turned on spell checker, then Oracle SES only gives suggestions for term search, phrase search, and proximity search. Spell checker does not give suggestions for terms or phrases in attribute shortcut search.

Alternate words expansion is turned off for searches that use advanced constructs like thesaurus-based search, proximity search, fuzzy search, phrase search, and compulsory exclusion search. See Table 3–1, " Search String Rules".

> **See Also:** "Alternate Words" on page 3-15

Auto Suggestions

Oracle SES can provide suggestion phrases for the query term typed in the search box, in the form of a list. Auto Suggestions help in reducing the keystrokes by narrowing down to the most appropriate query while entering a query term.

The suggestion phrases are retrieved from the query log that contains the most popular queries, that is, the queries which are searched very often by most of the users.

You can add and configure suggestion phrases in Oracle SES using Administration API. A suggestion phrase can be categorised into multiple *classifications*, and a classification can be assigned to multiple source groups.

The following example shows the relationship among suggestion phrases, classifications, and source groups:

Suggestion Keyword	Classification	Source Group
Secure Enterprise Search	Product	Oracle Products
Secure Backup	Product	Oracle Products
Smith	Name	People
Steve	Name	People
abc@oracle.com	Email	People
xyz@oracle.com	Email	People

To know the suggestion phrases that are configured in Oracle SES application, you can use the following URL:

```
http://ses_host_name:ses_port/search/query/AutoSuggestQuery?key=query_
term&sg=source_group&lang=language_code
```
where,

query_term is the term for which you want to know the suggestion phrases that are already configured in Oracle SES application.

source_group is the name of the source group for which the suggestion phrases are configured. This value is optional.

language_code is the two-letter language code adhering to the ISO 639-1 standard. The language code represents the language in which the suggestion phrases that you want to search are configured in Oracle SES application. For example, for English language use the language code of en.

> **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for information about the auto suggestions related Administration API objects - suggestion, autoSuggestion, classification, classificationMappings, and queryUIConfig - for creating, configuring, deleting, and exporting auto suggestions.

Linked Results

The results can include the following links:

Cached: The cached HTML version of the document.

Links: Pages that link to and from this document.

Source Group: Browse the source group.

Any links on top of the search text box are source groups. Clicking a source group restricts the search to that group.

Syntax Rules

Table 3–1 describes rules that apply to the search string. Text in square brackets represents characters entered into the search.

> **Note:** The use of otext:: syntax is not in supported in a search string in the query application as well as in the query Web Services API.

Table 3–1   Search String Rules

Rule	Description
Single term search	Entering one term finds documents that contain that term.
	For example, [Oracle] matches all documents that contain the word Oracle anywhere in that document.
	You can enter any two searchable items (including term, phrase, attribute shortcut, and proximity search) in a query with a white space separating them and the AND operator applies. The operator [&] also explicitly denotes an AND relationship.
	For example, [oracle text] and [oracle & text] both return documents containing oracle and text.

Table 3–1 (Cont.) Search String Rules

Rule	Description
Phrase search ["..."]	Put quotes around a set of words to find documents that contain that exact phrase.
	Oracle SES does not apply implicit stemming expansion to a query phrase, but it can apply explicit term expansion to terms in a phrase. All operators *except* term expansion operators in a phrase are not treated as valid operators but as normal special characters.
	For example, [oracle "RAC performance"] returns documents containing oracle and the phrase "RAC performance". Documents containing the stemming form "RAC performances" are not returned. (There is no implicit stemming expansion on either term.) The query ["sec*re search"] returns documents with the phrase "secure search". The query ["sec^re search"] returns documents with the phrase "sec re search".
Attribute shortcut search [*attribute_nume:attribute_value*]	Search on attributes with an attribute name, a colon (:), and then the value to be searched. Implicit stemming is applied to the attribute value term. You can specify operators as options. When no operator is specified, Oracle SES uses Contains for STRING attributes and Equals for NUMBER and DATE attributes.
	For example, [DocVersion:>1] returns documents that have number attribute Docversion where attribute value is larger than 1. The query [title:"oracle text"] returns documents with the phrase "oracle text" in the title attribute. The query [oracle | title:S*S] returns documents with the term oracle or SES in the title attribute. The query [title:^oracle] has the same effect as [title:oracle]. The contains [^] operator applies only to the STRING attribute.
	▪ Equals [=] returns documents with an attribute equaling the query with case-insensitivity. For example, [title:="oracle text"] returns documents whose title equals "oracle text". It applies to all three attributes.
	▪ Lessthan and narrower terms [<] return documents with an attribute value less than or earlier than the query value. For example, [DocVersion:<2] returns documents that have number attribute Docversion and where the attribute value is less than 2. They apply to all three attributes.
	▪ Lessthanequals [<=] applies to NUMBER and DATE attributes.
	▪ Greaterthan and broader terms [>] return documents with an attribute value greater than or later than the query value. They apply to all three attributes.
	▪ Greaterthanequals [>=] applies to NUMBER and DATE attributes. For example, [price:>=10] returns documents whose price attribute value is larger than or equal to 10. The query [lastmodifieddate:>=12/23/2006] returns documents whose lastmodifieddate attribute value is on or after December 23, 2006.
	See Also: "Searching on Date Attributes" on page 3-9 and "Advanced Search" on page 3-11

Table 3–1   (Cont.) Search String Rules

Rule	Description
Proximity search ["..."~]	A proximity search specifies the maximum distance within which multiple terms occur. A proximity search must have the search terms in double quotes. When the maximum spanning distance is not specified, Oracle SES applies a default window of 100 terms. The maximum number is 100. When a value larger than 100 is specified, Oracle SES treats it as 100.
	For example, ["ses performance"~10] returns documents with the terms SES and performance within any 10 terms spanning windows. The query ["ses performance"~] returns documents containing the terms SES and performance within any 100 terms spanning windows.
	Implicit stemming expansion is applied to each term in proximity search. Term expansion operators can be applied to terms in proximity search.
Fuzzy [...~] search	Put the operator (~) after a single term to return documents that contain terms similar to the query term.
	For example, [hallo~] returns documents containing term hello. The query [specifi*tion~] returns documents containing the term specification.
	Note: If a single term enclosed in double quotes is followed by ~, then the query is not a proximity search but a fuzzy search. The query ["parformance"~] returns documents containing the term performance.
Thesaurus-based search Synonym [~...] search Narrower term [<] search Broader term [>] search	Thesaurus-based operators require that a thesaurus be loaded into Oracle SES.
	Put the operator [~] at the beginning of a term to return documents that contain the original query term or a synonym for it. For example, [title:~"RAC"] returns documents with RAC or the synonym real application clusters in the title. A synonym relationship is symmetric: real application clusters is a synonym of RAC, and RAC is a synonym of real application clusters. In attribute search, it applies only to the STRING attribute.
	The query [<"Northern California"] returns documents with the thesaurus-defined narrower term San Francisco or the original phrase Northern California. The query [product:>chair] returns documents whose product attributes contain the broader term furniture or the original term chair. Broader and narrower terms are symmetric. Specifying that furniture is a broader term of chair also implicitly specifies that chair is a narrower term of furniture.
	See Also: "Thesaurus-Based Search" on page 3-14
OR [	] search	Use the OR [	]operator to connect any two searchable items.
	For example, [oracle	"RAC performance"~] returns documents with the term oracle or with the terms RAC and performance in any 100 terms spanning windows. The query [oracle	title:SES] returns documents with the term oracle or SES in the title attribute.
Grouping ( ) search	Use parentheses ( ) to group query components to change precedence of the binary logical operators AND and OR. The grouped query components must form a valid query. If the query string inside parentheses is not a valid query, then Oracle SES implicitly rewrites it to the closest valid query.
	For example, [(oracle | database) sales] returns documents containing sales *and* containing either oracle or database. The query [(oracle |) sales] returns documents containing oracle and sales. This is because [oracle |] is not a valid query.

Table 3–1   (Cont.) Search String Rules

Rule	Description
Wildcard matching [*] for multiple characters	Put the operator [*] in the middle or end of a term for wildcard matching. It can be applied multiple times in one term.
	For example, [ora*] finds documents that contain words beginning with ora, such as Oracle and orator. The query [title:a*e] returns documents with the title containing words such as apple or ape.
	Multiple character wildcard expansion could result in too many results. For example, [a*] could find too many results. Oracle SES throws an error to refine the queries.
	The wildcard operator [*] is ineffective with the escape character [\] just before it. For example [Pro*c].
	Wildcard matching cannot be used with Chinese or Japanese native characters.
Wildcard matching [?] for single characters	Put the operator (?) in middle or end of a term for wildcard matching for a single character. It can be applied multiple times in one term.
	For example, [orac?e] and [or?cl?] both return documents containing terms that replace ? with a single character, such as Oracle.
	The wildcard operator [?] is ineffective with the escape character [\] just before it.
	Wildcard matching cannot be used with Chinese or Japanese native characters.
Compulsory inclusion [+] search	Put the operator (+) at the beginning of any searchable item (including term, phrase, attribute, and proximity search) to require that the word be found in all matching documents. There should be no space between the [+] and the search term.
	For example, searching for [Oracle +Applications] only finds documents that contain the words Oracle and Applications.
	When compulsory inclusion search is used with the OR (	) operator, the compulsory inclusion operator does not have any effect. For example, searching for [text	+database] returns documents containing the term text or database.
Compulsory exclusion [-] search	Put the operator (-) at the beginning of any searchable item (including term, phrase, attribute, and proximity search) to require that the word not be found in all matching documents. It can be a single word or a phrase, but there should be no space between the [-] and the token.
	For example, [oracle -applications] returns documents containing oracle but *not* containing applications. The query [oracle -"application server"] returns documents containing oracle but not containing the phrase "application server". The query [oracle -title:oracle] returns documents containing oracle but with the title not containing oracle. The query [oracle -"application server"~] returns documents containing oracle but not containing application and server in any 100 terms spanning window.
	The compulsory exclusion query cannot be the only query. For example, the query [-oracle] raises an error. Also, the compulsory exclusion query cannot be connected with the OR [	] operator. For example, [oracle	-database] raises an error.

Table 3–1   (Cont.) Search String Rules

Rule	Description
Filetype search [filetype:*filetype*]	Use [filetype:*filetype*] after the search term to limit results to that particular file type. A search can have only one file type. No operator is allowed in file type shortcut search.
	For example, [documentation filetype:pdf] returns PDF format documents for the term documentation. The "filetype" shortcut must be lowercase, but the file type name is case-insensitive; that is, [documentation filetype:PDF] returns the same documents.
	The following file types are supported, with their corresponding mimetype:
	filetype string: mimetype
	ps: application/postscript
	ppt: application/vnd.ms-powerpoint, application/x-mspowerpoint, application/vnd.openxmlformats-officedocument.presentationml.presentation
	doc: application/msword, application/vnd.openxmlformats-officedocument.wordprocessingml.document
	xls: application/vnd.ms-excel, application/x-msexcel, application/ms-excel, application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
	txt: text/plain
	html: text/html
	htm: text/html
	pdf: application/pdf
	xml: text/xml
	rtf: application/rtf
Site search [site:*host*]	Use [site:*host*] after the search term to limit results to that particular site.
	For example, [site:www.oracle.com filetype:pdf] returns documents from www.oracle.com in PDF format. The "site" shortcut must be lowercase, but the host name is case-insensitive; that is, [site:www.Oracle.com filetype:pdf] returns the same documents.
	Oracle SES only supports exact host matching. The query [site:*.oracle.com] does not work.
Group search [sg:*source group*]	Use [sg:*source group*] to limit results to that particular source group. All other search restrictions are valid in a group search.
	For example, [sg:intranet] returns documents in the intranet source group. The "sg" shortcut must be lowercase, but the source group name is case-insensitive; that is, [sg:IntraNet] returns the same documents.
	In federated searches, the source group names are the source groups in the local (broker) node. If the local source groups contain federated sources, then Oracle SES translates the local source group name to the federated source group name by changing the query, which is then sent to federated source for results.

Browse Source Groups

Source groups are groups of sources that can be searched together. A source group consists of one or more sources, and a source can be assigned to multiple source groups. Source groups are defined on the Search - Source Groups page. Infosource nodes, or folders, are only generated for Web, e-mail, and OracleAS Portal source types.

On the basic Search page, users can browse source groups that the administrator created. Click a source group name to see the subgroups under it, or drill down further into the hierarchy by clicking a subgroup name.

To view all the documents under a particular group, click either **View All** or the number next to the source group name. You can also perform a restricted search in the source group from this page.

The source hierarchy lets end users limit search results based on document source type. The hierarchy is generated automatically during crawl time.

Searching on Date Attributes

Date attribute values on the result list are shown in Greenwich Mean Time (GMT). For example, when you crawl a document on your local computer with a last modified date value of "Sep 13 2007 20:30:00 PDT", the Oracle SES crawler converts this date value to the corresponding GMT date value, which is "Sep 14 2007 3:30:00 GMT". These two values represent the same moment in time, but Oracle SES only displays the date (not the time or time zone). Therefore, the last modified date displayed in the result list is Sep 14 2007 and not Sep 13 2007.

To search on the `lastModifiedDate` attribute, use the GMT date value. In the previous example, you would search on `lastModifiedDate=09/14/2004`. The query `lastModifiedDate=09/13/2004` does not return the document. Note that you must enter the date using the format *mm/dd/yyyy*.

To display the last modified date in your local time zone:

1. Log in to the Oracle SES Administration GUI.

2. On the Global Settings page, click Query UI Configuration.

3. Select **Yes** for Convert to Local Timezone field.

The browser picks up your local time zone and `lastModifiedDate` is converted to your local time zone before displaying the search results.

Submit URL

The URL submission feature lets users submit URLs to be crawled and indexed. These URLs are added to the seed URL list for a particular source and are included in the crawler search space.

This feature is disabled on the Search page if it is disabled on the Global Settings - Query Configuration page or if you have not created a Web source.

To allow users to submit URLs:

1. Log in to the Oracle SES Administration GUI.

2. On the Global Settings page, select **Query Configuration**.

3. Select **Allow URL Submission**.

4. Select the Web source to which the submitted URLs are added.

5. Click **Apply**.

To examine the submitted URLs before they are indexed by the crawler:

1. On the home page, select the Schedules secondary tab.

2. Click the **Edit** icon of a schedule.

3. Under Update Crawling Mode on the Edit Schedule page, select **Examine URLs Before Indexing**.

4. Click **Update Crawling Mode**.

Configuring the Display Options on the Search Page

You can configure how search options are displayed on the Search page. You can choose to display the Attribute Filters link or the Advanced Search link next to the basic search option.

This is done by specifying the following setting at the beginning of each of the four Freemarker template files: `query.ftl`, `results.flt`, `noresults.ftl`, and `error.ftl`.

```
<#assign advSearchOpt = "configuration_setting">
```

For example,

```
<#assign advSearchOpt = "filters">
```

The configuration options are:

- `filters` — Displays the Attribute Filters link next to the Search button. Using this link, you can access the attributes filter table. This is the default display setting.

- `link` — Displays the Advanced Search link next to the Search button. Using this link, you can access the advanced search page.

- `both` — Displays the Attribute Filters link next to the Search button. The `More Advanced` link at the bottom of the attribute filters table takes you to the Advanced Search page.

> **See Also:** Chapter 10, "Customizing the Search User Interface" for more information about configuring Freemarker template files

Attribute Filters

Attribute filters enable you to refine your search query by specifying attribute values of documents. Oracle SES includes default attributes for every instance, such as title, description, and keywords. For example, the Author attribute is mapped to the From header in e-mail documents and the Author metatag in HTML documents. Search administrators can also create custom attributes.

Using the Attributes Filter Table

The attributes filter table provides several user friendly features that makes it easier for you to create a search query based on attributes. The features include auto-complete for attribute names that match the user's input, and a calendar picker that is displayed when the user selects a date attribute. In the Attribute Filters table, you can specify any number of attribute filters by specifying a boolean condition followed by "any" or "all". Attribute filter rows may be added or removed by clicking on the icons to the right of each row.

Auto Complete for Attribute Names

To assist the user in selecting search attributes, auto complete is enabled for the attribute name. As you begin entering the attribute name, the attributes with the matching characters are displayed in a list and you can select from this list. Additionally, the list displays the data type next to the attribute name.

If a source group or multiple source groups are selected, then this list of search attributes is restricted to those attributes that are included in any of the selected source groups.

Attribute Type-Specific Operators

The choice of operator is specific to the data type of the selected attribute. For example, STRING attributes allow the Equals and Contains operators. NUMBER and DATE attributes allow Equals, Less than, Less than or equals, Greater than, and Greater than or equals operators. Also note that the choice of available operators depends on the attribute itself. For example, with the Tag attribute, you can use only the Equals operator.

Calendar Date Picker for Date Attributes

When a date attribute is selected and a user clicks or tabs into the attribute value input box for the date attribute, a calendar control automatically appears to allow the user to easily select a date. In addition to providing a visual interface, this also aids the user in entering the correct date format for Oracle SES, which is currently MM/DD/YYYY.

Link Label in Collapsed State

When the attribute filters table is collapsed/hidden, the Attribute Filters link indicates the number of valid attribute filters.

Although there may be many more rows in the attribute filter table, this number only counts rows that have both a valid attribute name and a non-empty value. Incomplete rows are ignored if the query is submitted, so this number is an accurate count of the filters in the search query.

Display Query for Advanced Search

When a search is performed with the attributes specified in the attributes filters table, the equivalent "advanced search" query with attribute shortcuts is constructed to display to the user various statistics like the title, hit statistics, and bottom query box. This shows the user what the equivalent query would be if entered solely in the query box.

See "Search by Attribute" on page 3-13 for more information.

Advanced Search

Oracle SES 11*g* provides a feature-rich set of advanced search options for the end user. You can target your search results more accurately using the Advanced Search page, which is shown in Figure 3–3. Advanced Search lets you refine searches in the following ways:

- Search Documents in a Specific Language

- Search in Specific Source Groups

- Search by Attribute

Figure 3–3   Advanced Search Page

Search Documents in a Specific Language

Oracle SES can search documents in different languages. Specifying a language restricts searches to documents that are written in that language. Select a language in the Language box.

The following are possible internationalization issues with supported operators:

- Proximity Search ["secure search"~10]: The term distance definition could be different for non-whitespace delimited languages, such as Japanese. The behavior of proximity search for those languages could be different.

- Implicit stemming: This is available in English, German, French, Spanish, Italian, Dutch.

- Wildcard search [feat*e] or [featu?e]: The term definition is different for non-whitespace-delimited languages, such as, Chinese and Japanese. With Oracle SES default index preference configuration (CHINESE_VGRAM_LEXER and JAPANESE_VGRAM_LEXER), the behavior of wildcard expansion for Chinese and Japanese languages is not applicable, and hence wildcard operator cannot be used with Chinese and Japanese characters.

- Fuzzy expansion [hallo~]: This is available in English, German, French, Chinese, Japanese, Spanish, Italian, and Dutch.

Search in Specific Source Groups

If one or more source groups are defined in Oracle SES, then corresponding check boxes appear when you click the Browse button in the query application. You can limit your search to specific source groups by selecting the corresponding source group check boxes. If no source group is selected, then the documents from the configured default source group are searched. The default source group is specified in the **Default Source Group** field of the **Query UI Configuration** page of the Administration GUI. Generally, the default source group name is **All (System)**, and in the query application, it is shown as **All** source group link. When **All** source group link is clicked, the query application crawls all the sources during the search operation.

Customizing All Source Group

If you want to customize the functionality of the default **All (System)** source group provided by Oracle SES, then you can define a custom **All** source group to replace the default **All (System)** source group. The custom **All** source group is like any other source group, but you need to explicitly add all the sources to it, and it should always be updated with the new sources whenever they are created in Oracle SES.

The following are the steps to replace the default **All (System)** source group with the custom **All** source group:

1. Modify skinBundle to hide the default **All** source group link shown in the query application. The following are the steps to achieve this:

   a. Create a new skinBundle directory by copying to it the skinBundle template files `query.ftl` and `inc_logo_querybox.ftl`. These template files are stored in the directory:

      `wls_domain_home/ses_domain_name/servers/search_server1/tmp/_WL_user/search_`
      `query/curkae/war/WEB-INF/templates`

   b. Edit the `query.ftl` file present in the new skinBundle directory by opening it in a text editor and adding the argument `showDefaultAll=false` to the `ses.source_group_tabs` entry as shown in the following example:

      ```
      <@ses.source_group_tabs formName="${searchFormName}"
        suppressGroups=suppressSrcGroups![] showDefaultAll=false/>
      ```

   c. Create the skinBundle XML file.

   d. Create the skinBundle mapping file.

   e. Reload the skinBundle by executing the following commands:

      ```
      searchadmin -p password -c http://host:port/search/api/admin/AdminService
      create skinBundle -n skinName -i skinBundleXMLFileName.xml -h
      attachmentFileList.lst

      searchadmin -p password -c http://host:port/search/api/admin/AdminService
      activate skinBundle -n skinName
      ```

2. Create a new source group named **All** in the Administration GUI by clicking the **Create** button in the **Source Group** sub-tab under the **Search** tab, and add all the existing sources to it. You should always update the **All** source group with the new sources whenever they are created in Oracle SES.

   > **See Also:** "Using Skin Bundles to Customize the Search GUI" on page 10-13

Search by Attribute

Similar to the attributes filter table, the Attribute Selection panel enables you to specify attribute values for a search. Date format must be entered in the MM/DD/YYYY format. Click **Add more attributes** to enter more than four search attribute values. Table 3–2 provides the list of operators supported for each data type.

Table 3–2   Oracle SES Attribute Search Operators and Types of Attributes

Attribute	Contains (^)	Equals (=)	Synonym (~)	Lessthan, Narrower terms (<)	Lessthan equals (<=)	Greaterthan, Broader terms (>)	Greaterthanequals (>=)
String	Yes	Yes	No	No	No	No	No
Number	No	Yes	No	Yes	Yes	Yes	Yes

Table 3–2   (Cont.) Oracle SES Attribute Search Operators and Types of Attributes

Attribute	Contains (^)	Equals (=)	Synonym (~)	Lessthan, Narrower terms (<)	Lessthan equals (<=)	Greaterthan, Broader terms (>)	Greatertha nequals (>=)
Date	No	Yes	No	Yes	Yes	Yes	Yes

The `Equals` operator returns a hit only if the attribute value that you enter *exactly* matches the attribute value in the document. The `Contains` operator returns a hit if the attribute value you enter matches *any* of the tokens in the attribute values in the document. The token must be an exact match; partial matches are not returned.

For example, suppose that the following four documents are indexed:

Document	Author	Number of Tokens
doc1	"scott"	
doc2	"scott tiger"	2 (scott, tiger)
doc3	"scottm tiger"	2 (scottm, tiger)
doc4	"scott.tiger@xyz.com"	4 (scott, tiger, xyz, com)

An attribute restricted search for "author equals scott" returns only doc1. But an attribute restricted search for "author contains scott" returns doc1, doc2, and doc4. There is no hit for doc3 because `scott` is only a partial match for `scottm`.

Thesaurus-Based Search

A thesaurus is a list of terms or phrases with relationships specified among them, such as a synonym, a broader term, and a narrower term. When a user issues a search query, Oracle SES can expand the search results to include matches for the related terms. Users can include the thesaurus operators in their queries:

- ~ Synonyms
- \> Broader terms
- < Narrower terms

If no thesaurus is loaded or if the specified term or phrase cannot be found in the loaded thesaurus, then query expansion is not possible. Oracle SES only returns documents containing the original term.

You can apply thesaurus-based expansion operators to attributes in an attribute search. Because an attribute value is either a term or a phrase, the expansion has the same effect on both, except that the expansion in attribute shortcut search is restricted to attribute value search.

Including Related Terms from a Thesaurus

You can expand the search results by providing a thesaurus. A thesaurus is a list of terms or phrases with relationships specified among them, such as a synonym, a broader term, and a narrower term. When a user issues a search query, Oracle SES can expand the search results to include matches for the related terms.

A thesaurus contains domain-specific knowledge. You can build a thesaurus, buy an industrial-specific thesaurus, or use utilities to extract a thesaurus from a specific

corpus of documents. The thesaurus must be compliant with both the ISO-2788 and ANSI Z39.19(1993) standards.

Only one thesaurus can be loaded at a time, and it must be named DEFAULT. If no thesaurus is loaded or if the specified term or phrase cannot be found in the loaded thesaurus, then thesaurus-based query expansion is not possible. Oracle SES only returns documents containing the original term or phrase. The default expansion level is 1.

> **See Also:**
> - *Oracle Secure Enterprise Search Administration API Guide*
> - *Oracle Text Application Developer's Guide*

Alternate Words

Oracle SES maintains an alternate word list containing word pairs. The two words in the word pair can be used alternatively. The semantic similarity between the two words is higher than that between two synonyms.

Oracle SES uses alternate words in the following places:

- To provide a suggestion for a query. For example, the query [RAC] can result in the suggestion: "did you mean 'real application clusters'?"
- To provide an implicit expansion based on alternate words. For example, the query [RAC] returns documents containing RAC or real application clusters.

Oracle SES provides an option for each alternate word pair to do implicit expansion for this pair.

Suggesting Alternate Words

Oracle SES can maintain an alternate word list containing word pairs. The two words in the word pair can be used alternatively. The semantic similarity between the two words is higher than that between two synonyms. Oracle SES uses the list to provide suggestions or to expand the search results.

> **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for more information about configuring alternate words using the `altWord` object.

Restricted Search and the Browse Popup

On the Basic search page, users can browse source groups that the administrator created by clicking the **Browse** link next to the **Search** box. This action displays a Browse popup window containing a tree view of the browse information hierarchy. Users can click an expand icon (>) to see the **infosource** nodes under it, or drill down further by clicking additional expand icons.

Clicking the **View All** link or the document count number of a browse node refreshes the entire page to show the set of all documents within that infosource node. Clicking the node label causes the list of source groups above the **Search** box to be replaced with the message "Search within: *infosource node name*".

The infosource node in the browse tree, along with its subtree, are highlighted to indicate which node is selected to search within.

The search result set is not immediately replaced. Instead, users must click **Search** to perform a restricted search within the selected infosource node. Only one infosource node may be selected at a time for "search-within."

To restrict search to a set of top-level source groups (rather than an infosource node), select multiple source groups within the Browse Tree popup. The source group has a check mark next to it, and the list of selected source group names are displayed above the **Search** box. Again, users must click **Search** to perform a restricted search within the selected source groups.

Personalized Search using Preference Profiles

You can configure multiple personal profiles in the query application for controlling the behavior of query results according to your requirement.

A preference profile is a group of configuration settings specific to a user, such as, number of documents to display, source groups to display, and the search filter criteria for specific source groups as well as for all the source groups.

You can also share your preference profiles with other users, and can subscribe to or copy the preference profiles created by other users, only if the other users have shared their preference profiles with you.

You can configure the preference profiles by clicking the **Preferences** link in the query application's default query page. The **Preferences** link is displayed only when you are logged in to the query application. The Preference Profiles page enables you to create, update, and delete your preference profiles.

> **Note:** The preference profile configuration takes precedence over the Query UI configuration specified in the Administration GUI.

See Also: Oracle SES Query Application Online Help

4

Understanding Crawling

This chapter discusses the Oracle SES crawler. It contains the following topics:

- Overview of the Oracle Secure Enterprise Search Crawler
- Overview of Crawler Settings
- Overview of XML Connector Framework
- Overview of Document Service
- Configuring Support for Image Metadata
- Overview of Attributes
- Understanding the Default Crawling Process
- Understanding the Push Type Crawling Process
- Configuring the Crawler Settings using Administration API
- Configuring the Crawler Log
- Monitoring the Crawling Process
- Parallel Query Indexing

> **See Also:**
>
> - "Tuning Crawl Performance" on page 12-5 and "Tuning Search Performance" on page 12-13
> - *Oracle SES Administration Tutorial* at `http://www.oracle.com/pls/topic/lookup?ctx=E35215-01&id=tut_ses_administration`

Overview of the Oracle Secure Enterprise Search Crawler

The Oracle Secure Enterprise Search (Oracle SES) crawler is a Java process activated by a set schedule. When activated, the crawler spawns processor threads that fetch documents from **sources**. The crawler caches the documents, and when the cache reaches the maximum batch size of 250 MB, the crawler indexes the cached files. This index is used for searching.

The document cache, called Secure Cache, is stored in Oracle Database in a compressed SecureFile LOB. Oracle Database provides excellent security and compact storage.

In the Oracle SES Administration GUI, you can create schedules with one or more sources attached to them. Schedules define the frequency at which the Oracle SES index is kept up to date with existing information in the associated sources.

See "Understanding the Default Crawling Process" on page 4-27 for more detailed information about the crawling process.

Crawler URL Queue

In the process of crawling, the crawler maintains a list of URLs of the discovered documents that are fetched and indexed in an internal URL queue. The queue is persistently stored, so that crawls can be resumed after the Oracle SES instance is restarted.

Understanding Access URLs and Display URLs

A display URL is a URL string used for search result display. This is the URL used when users click the search result link. An access URL is an optional URL string used by the crawler for crawling and indexing. If it does not exist, then the crawler uses the display URL for crawling and indexing. If it does exist, then it is used by the crawler instead of the display URL for crawling. For regular Web crawling, only display URLs are available. But in some situations, the crawler needs an access URL for crawling the internal site while keeping a display URL for the external use. For every internal URL, there is an external mirrored URL.

For example, for file sources with display URLs, end users can access the original document with the HTTP or HTTPS protocols. These provide the appropriate authentication and personalization and result in better user experience.

Display URLs can be provided using the URL Rewriter API. Or, they can be generated by specifying the mapping between the prefix of the original file URL and the prefix of the display URL. Oracle SES replaces the prefix of the file URL with the prefix of the display URL.

For example, if the file URL is

```
file://localhost/home/operation/doc/file.doc
```

and the display URL is

```
https://webhost/client/doc/file.doc
```

then specify the file URL prefix as

```
file://localhost/home/operation
```

and the display URL prefix as

```
https://webhost/client
```

Modifying the Crawler Parameters

You can alter the crawler's operating parameters at two levels:

- At the global level for all sources
- At the source level for a particular defined source

Global parameters include the default values for language, crawling depth, and other crawling parameters, and the settings that control the crawler log and cache.

To configure the crawler:

1. Click the **Global Settings** tab.

2. Under Sources, click **Crawler Configuration**.

3. Make the desired changes on the Crawler Configuration page. Click **Help** for more information about the configuration settings.

4. Click **Apply**.

To configure the crawling parameters for a specific source:

1. On the home page, click the **Sources** secondary tab to see a list of sources you have created.

2. Click the edit icon for the source whose crawler you want to configure, to display the Edit Source page.

3. Click the **Crawling Parameters** subtab.

4. Make the desired changes. Click **Help** for more information about the crawling parameters.

5. Click **Apply**.

Note that the parameter values for a particular source can override the default values set at the global level. For example, for Web sources, Oracle SES sets a default crawling depth of 2, irrespective of the crawling depth you set at the global level.

Also note that some parameters are specific to a particular source type. For example, Web sources include parameters for HTTP cookies.

Overview of Crawler Settings

This section describes crawler settings and other mechanisms to control the scope of Web crawling:

- Crawling Mode
- URL Boundary Rules
- Document Types
- Crawling Depth
- Robots Exclusion
- Sitemap Crawl
- Index Dynamic Pages
- URL Rewriter API
- Title Fallback
- Character Set Detection

> **See Also:** "Tuning Crawl Performance" on page 12-5 for more detailed information on these settings and other issues affecting crawl performance

Crawling Mode

For initial planning purposes, you might want the crawler to collect URLs without indexing. After crawling is finished, examine the document URLs and status, remove unwanted documents, and start indexing. The crawling mode is set on the Home - Schedules - Edit Schedules page.

> **See Also:** Appendix B, "URL Crawler Status Codes"

> **Note:** If you are using a custom crawler created with the Crawler Plug-in API, then the crawling mode set here does not apply. The implemented plug-in controls the crawling mode.

These are the crawling mode options:

- **Automatically Accept All URLs for Indexing**: This crawls and indexes all URLs in the source. For Web sources, it also extracts and indexes any links found in those URLs. If the URL has been crawled before, then it is reindexed only if it has changed.

- **Examine URLs Before Indexing**: This crawls but does not index any URLs in the source. It also crawls any links found in those URLs.

- **Index Only**: This crawls and indexes all URLs in the source. It does not extract any links from those URLs. In general, select this option for a source that has been crawled previously under "Examine URLs Before Indexing".

URL Boundary Rules

URL boundary rules limit the crawling space. When boundary rules are added, the crawler is restricted to URLs that match the indicated rules. The order in which rules are specified has no impact, but exclusion rules always override inclusion rules.

This is set on the Home - Sources - Boundary Rules page.

Inclusion Rules

Specify an inclusion rule that a URL contain, start with, or end with a term. Use an asterisk (*) to represents a wildcard. For example, `www.*.example.com`. Simple inclusion rules are case-insensitive. For case-sensitivity, use regular expression rules.

An inclusion rule ending with example.com limits the search to URLs ending with the string `example.com`. Anything ending with `example.com` is crawled, but `http://www.example.com.tw` is not crawled.

If the URL Submission functionality is enabled on the Global Settings - Query Configuration page, then URLs that are submitted by end users are added to the inclusion rules list. You can delete URLs that you do not want to index.

Oracle SES supports the regular expression syntax used in the JDK Pattern class (`java.util.regex.Pattern`). Regular expression rules use special characters. The following is a summary of some basic regular expression constructs.

- A caret (^) denotes the beginning of a URL and a dollar sign ($) denotes the end of a URL.

- A period (.) matches any one character.

- A question mark (?) matches zero or one occurrence of the character that it follows.

- An asterisk (*) matches zero or more occurrences of the pattern that it follows. You can use an asterisk in the starts with, ends with, and contains rules.

- A backslash (\) escapes any special characters, such as periods (\.), question marks (\?), or asterisks (*).

> **See Also:** http://docs.oracle.com/javase/6/docs for a complete description about the JDK documentation

Exclusion Rules

You can specify an exclusion rule that a URL contains, starts with, or ends with a term. For case-sensitivity, use regular expression rules, as described in the Inclusion Rules section.

An exclusion of `uk.example.com` prevents the crawling of Example hosts in the United Kingdom.

Default Exclusion Rules

The crawler contains a default exclusion rule to exclude non-textual files. The following file extensions are included in the default exclusion rule.

- **Image**: jpg, gif, tif, bmp, png

- **Audio**: wav, mp3, wma

- **Video**: avi, mpg, mpeg, wmv

- **Binary**: bin, exe, so, dll, iso, jar, war, ear, tar, wmv, scm, cab, dmp

> **Note:** Only the file name is indexed when crawling multimedia files, unless the file is crawled using a crawler plug-in that provides a richer set of attributes, such as the Image Document Service plug-in.

Example Using a Regular Expression

The following example uses several regular expression constructs that are not described earlier, including range quantifiers, non-grouping parentheses, and mode switches. For a complete description, see the Sun Microsystems Java documentation.

To crawl only HTTPS URLs in the `example.com` and `examplecorp.com` domains, and to exclude files ending in .doc and .ppt:

- Inclusion: URL regular expression

  `^https://.*\.example(?:corp){0,1}\.com`

- Exclusion: URL regular expression `(?i:\.doc|\.ppt)$`

Document Types

Oracle SES uses Oracle Text to convert binary documents to HTML. See "Appendix B" of *Oracle Text Reference* for the complete list of document types supported by Oracle SES.

The following table describes the document types that are processed by default by the crawler for various source types:

Source Types	Document Types (MIME Types) Processed by Default
Oracle Content Server, Oracle E-Business Suite, Oracle Fusion, Oracle Web Center, PeopleSoft, and Siebel 8.	All the supported document types are processed.

Source Types	Document Types (MIME Types) Processed by Default
Web, File, Microsoft NTFS, WebDAV, EMC Documentum Content Server, Table, Database, E-mail, Mailing List, Microsoft Exchange, Oracle Collaboration Suite E-mail, Oracle Calendar, OracleAS Portal, Microsoft Sharepoint 2007, Lotus Notes, Oracle Content Database, Oracle Content Database (JDBC), Siebel 7.8, and Siebel 7.8 (Public).	text/html, application/x-msexcel, application/x-mspowerpoint, application/msword, application/pdf, text/plain

Configuring Document Types to Process for a Source using Administration GUI

To configure the document types to process for a source using Administration GUI:

1. On the Administration GUI home page, click the **Sources** tab.

2. Choose a source from the list of sources and select Edit to display the Customize Source page.

3. Select the **Document Types** sub-tab.

   The listed document types are supported for the source type.

4. Move the types to process to the Processed list and the others to the Not Processed list.

5. Click **Apply**.

Configuring Document Types to Process for a Source using Administration API

You can use the Administration API object `globalDocumentTypes` to specify the document types to process for a source. Refer to *Oracle Secure Enterprise Search Administration API Guide* for more information.

> **Note:**
>
> - For graphics format files (JPEG, JPEG 2000, GIF, TIFF, **DICOM**), only the file name is searchable. The crawler does not extract any metadata from graphics files or make any attempt to convert graphical text into indexable text, unless you enable a document service plug-in. See "Configuring Support for Image Metadata" on page 4-16.
>
> - If a zip file or an LHA file contain more than 1000 files in it, then the crawler ignores that file and stores the related error message in the log file.

Crawling Depth

Crawling depth is the number of levels to crawl Web and file sources. A Web document can contain links to other Web documents, which can contain more links. Specify the maximum number of nested links for the crawler to follow. Crawling depth starts at 0; that is, if you specify 1, then the crawler gathers the starting (seed) URL plus any document that is linked directly from the starting URL. For file crawling, this is the number of directory levels from the starting URL.

Set the crawling depth on the Home - Sources - Customize Web Source - Crawling Parameters page.

Robots Exclusion

You can control which parts of your sites can be visited by robots. If robots exclusion is enabled (default), then the Web crawler traverses the pages based on the access policy specified in the Web server robots.txt file. The crawler also respects the page-level robot exclusion specified in HTML metatags.

For example, when a robot visits `http://www.example.com/`, it checks for `http://www.example.com/robots.txt`. If it finds it, then the crawler checks to see if it is allowed to retrieve the document. If you own the Web sites, then you can disable robots exclusions. However, when crawling other Web sites, always comply with robots.txt by enabling robots exclusion.

Set the robots parameter on the Home - Sources - Customize Web Source - Crawling Parameters page.

Sitemap Crawl

Sitemap crawl provides an efficient way to crawl Web sources. A Sitemap is a file that generally resides under Web site's root directory and contains a list of URLs that a search engine or a crawler is allowed to crawl. Thus, only the relevant URLs on a Web site are crawled by the crawler.

> **Note:** The Sitemap crawl feature is disabled by default in Oracle SES. You can enable and configure the Sitemap feature in Oracle SES using Administration API. Once the Sitemap crawl feature is enabled, then it cannot be disabled, that is, Sitemap information is then always processed by the crawler.

The crawler can auto-detect a Sitmap file in the Web sites. The crawler searches for the Sitemap information in the `robots.txt` file under the Web site's root directory. If the `robots.txt` file does not contain the Sitemap information, then the crawler searches for the `sitemap.xml` file under the Web site's root directory. A Sitemap XML location URL can refer to a Sitemap file, a Sitemap index file, or a zip file (.gz) of these two file types.

> **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for configuring Sitemap in Oracle SES using the Administration API's `sitemap` element of the `source` object of the Web type source.

Index Dynamic Pages

By default, Oracle SES processes dynamic pages. Dynamic pages are generally served from a database application and have a URL that contains a question mark (?). Oracle SES identifies URLs with question marks as dynamic pages.

Some dynamic pages appear as multiple search results for the same page, and you might not want them all indexed. Other dynamic pages are each different and must be indexed. You must distinguish between these two kinds of dynamic pages. In general, dynamic pages that only change in menu expansion without affecting its contents should not be indexed.

Consider the following three URLs:

```
http://example.com/aboutit/network/npe/standards/naming_convention.html
```

```
http://example.com/aboutit/network/npe/standards/naming_convention.html?nsdnv=14z1
```

```
http://example.com/aboutit/network/npe/standards/naming_convention.html?nsdnv=14
```

The question marks (?) in two URLs indicate that the rest of the strings are input parameters. The three results are essentially the same page with different side menu expansion. Ideally, the search yields only one result:

```
http://example.com/aboutit/network/npe/standards/naming_convention.html
```

> **Note:** The crawler cannot crawl and index dynamic Web pages written in Javascript.

Set the dynamic pages parameter on the Home - Sources - Crawling Parameters page.

URL Rewriter API

The URL Rewriter is a user-supplied Java module for implementing the Oracle SES `UrlRewriter` interface. The crawler uses it to filter or rewrite extracted URL links before they are put into the URL queue. The API enables ultimate control over which links extracted from a Web page are allowed and which ones should be discarded.

URL filtering removes unwanted links, and URL rewriting transforms the URL link. This transformation is necessary when access URLs are used and alternate display URLs must be presented to the user in the search results.

Set the URL rewriter on the Home - Sources - Crawling Parameters page.

> **See Also:**
>
> - "URL Rewriter API" on page 13-45
> - *Oracle Secure Enterprise Search Java API Reference*

Title Fallback

You can override a default document title with a meaningful title if the default title is irrelevant. For example, suppose that the result list shows numerous documents with the title "Daily Memo". The documents had been created with the same template file, but the document properties had not been changed. Overriding this title in Oracle SES can help users better understand their search results.

Any source type can use title fallback. Oracle SES uses different logic for each document type to determine which fallback title to use. For example, for HTML documents, Oracle SES looks for the first heading, such as `<h1>`. For Microsoft Word documents, Oracle SES looks for text with the largest font.

If the default title was collected in the initial crawl, then the fallback title is only used after the document is reindexed during a re-crawl. This means if there is no change to the document, then you must force the change by setting the re-crawl policy to **Process All Documents** on the Home - Schedules - Edit Schedule page.

You can override the default document title with a meaningful title by using the `badTitle` element of the `crawlerSettings` object of Administration API. You can specify multiple bad titles. For example:

```
<search:crawlerSettings>
   ...
   <search:badTitles>
      <search:badTitle>PowerPoint Presentation</search:badTitle>
      <search:badTitle>Slide 1</search:badTitle>
    </search:badTitles>
```

```
   ...
</search:crawlerSettings>
```
This functionality is not provided in Administration GUI.

Special Considerations with Title Fallback

- With Microsoft Office documents:

  - Font sizes 14 and 16 in Microsoft Word correspond to normalized font sizes 4 and 5 (respectively) in converted HTML. The Oracle SES crawler only picks up strings with normalized font size greater than 4 as the fallback title.

  - Titles should contain more than five characters.

- When a title is null, Oracle SES automatically indexes the fallback title for all binary documents (for example, .doc, .ppt, .pdf). For HTML and text documents, Oracle SES does *not* automatically index the fallback title. The replaced title on HTML or text documents cannot be searched with the title attribute on the Advanced Search page.

  You can turn on indexing for HTML and text documents by using Administration API's `indexNullTitleFallback` element of the `source` object for Web type and File type source.

  > **See Also:** *Oracle Secure Enterprise Search Administration API Guide*

Character Set Detection

This feature enables the crawler to automatically detect character set information for HTML, plain text, and XML files. Character set detection allows the crawler to properly cache files during crawls, index text, and display files for queries. This is important when crawling multibyte files (such as files in Japanese or Chinese).

You can enable automatic character set detection by using the `charsetDetection` element of the `crawlerSettings` object of Administration API. For example:

```
<search:crawlerSettings>
   ...
   <search:charsetDetection>true</search:charsetDetection>
   ...
</search:crawlerSettings>
```
This functionality is not provided in Administration GUI.

Special Considerations with Automatic Character Set Detection

- To crawl XML files for a source, be sure to add XML to the list of processed document types on the Home - Source - Document Types page. XML files are currently treated as HTML format, and detection for XML files may not be as accurate as for other file formats.

Language Detection

With multibyte files, besides turning on character set detection, be sure to set the **Default Language** parameter. For example, if the files are all in Japanese, select Japanese as the default language for that source. If automatic language detection is disabled, or if the crawler cannot determine the document language, then the crawler assumes that the document is written in the default language. This default language is used only if the crawler cannot determine the document language during crawling.

If your files are in multiple languages, then turn on the **Enable Language Detection** parameter. Not all documents retrieved by the crawler specify the language. For

documents with no language specification, the crawler attempts to automatically detect language. The language recognizer is trained statistically using trigram data from documents in various languages (for instance, Danish, Dutch, English, French, German, Italian, Portuguese, and Spanish). It starts with the hypothesis that the given document does not belong to any language and ultimately refutes this hypothesis for a particular language where possible. It operates on Latin-1 alphabet and any language with a deterministic Unicode range of characters (like Chinese, Japanese, Korean, and so on).

The crawler determines the language code by checking the HTTP header content-language or the LANGUAGE column, if it is a table source. If it cannot determine the language, then it takes the following steps:

- If the language recognizer is not available or if it cannot determine a language code, then the default language code is used.

- If the language recognizer is available, then the output from the recognizer is used.

- Oracle SES uses different lexers for space-delimited languages (such as English), Chinese, Japanese, and Korean. See the lexer object description in the *Oracle Secure Enterprise Search Administration API Guide*.

The **Default Language** and the **Enable Language Detection** parameters are on the Global Settings - Crawler Configuration page (globally) and also the Home - Sources - Crawling Parameters page (for each source).

> **Note:** For file sources, the individual source setting for **Enable Language Detection** remains false regardless of the global setting. In most cases, the language for a file source should be the same, and set from, the **Default Language** setting.

Deleting the Secure Cache

You can manage the Secure Cache either on the global level or at the data source level. The data source configuration supersedes the global configuration.

The cache is preserved by default and supports the Cached link feature in the search result page. If you do not use the **Cache** link, then you can delete the cache, either for specific sources or globally for all of them. Without a cache, the **Cached** link in a search result page returns a File not found error.

To delete the cache for all sources:

1. Select the **Global Settings** tab in the Oracle SES Administration GUI.

2. Choose **Crawler Configuration**.

3. Set **Preserve Document Cache** to **No**.

4. Click **Delete Cache Now** to remove the cache from all sources, except any that are currently active under an executing schedule. The cache is deleted in the background, and you do not have to wait for it to complete.

5. Click **Apply**.

To delete the cache for an individual source:

1. Select the **Sources** secondary tab on the home page.

2. Click **Edit** for the source.

3. Click the **Crawling Parameters** subtab.

4. Set **Preserve Document Cache** to **No**.

5. Click **Apply**.

Overview of XML Connector Framework

Oracle SES provides an XML connector framework to crawl any repository that provides an XML interface to its contents. The connectors for Oracle Content Server, Oracle E-Business Suite 12, and Siebel 8 use this framework.

Every document in a repository is known as an **item**. An item contains information about the document, such as author, access URL, last modified date, security information, status, and contents.

A set of items is known as a **feed** or **channel**. To crawl a repository, an XML document must be generated for each feed. Each feed is associated with information such as feed name, type of the feed, and number of items.

To crawl a repository with the XML connector, place data feeds in a location accessible to Oracle SES over one of these protocols: HTTP, FTP, or File. Then generate an XML Configuration File that contains information such as feed location and feed type. Create a source with a source type that is based on this XML connector and trigger the crawl from Oracle SES to crawl the feeds.

There are two types of feeds:

- **Control feed**: Individual feeds can be located anywhere, and a single control file is generated with links to the feeds. This control file is input to the connector through the configuration file. A link in control feed can point to another control feed. Control feed is useful when data feeds are distributed over many locations or when the data feeds are accessed over diverse protocols such as FTP and file.

- **Directory feed**: All feeds are placed in a directory, and this directory is input to the connector through the configuration file. Directory feed is useful when the data feeds are available in a single directory.

Guidelines for the target repository generating the XML feeds:

- XML feeds are generated by the target repository, and each file system has a limit on how many files it can hold. For directory feeds, the number of documents in each directory should be less than 10,000. There are two considerations:

  - **Feed files**: The number of items in each feed file should be set such that the total number of feed files in the feed directory is kept under 10,000.

  - **Content files**: If the feed files specify content through attachment links and the targets of these links are stored in the file system, then ensure that the targets are distributed in multiple directories so that the total number of files in each directory is kept under 10,000.

- When feeds are generated real-time over HTTP, ensure that the component generating the feeds is sensitive to time out issues of feed requests. The feed served as the response for every request should be made available within this time out interval; otherwise, the request from Oracle SES times out. The request is retried as many times as specified while setting up the source in Oracle SES. If all these attempts fail, then the crawler ignores this feed and proceeds with the next feed.

See Also:

- Appendix A, "XML Connector Examples and Schemas"
- "Setting Up Oracle Content Server Sources" on page 7-25
- "Setting Up Oracle E-Business Suite Sources" on page 9-10
- "Setting Up Siebel 8 Sources" on page 9-30

Example Using XML Connector

The courses in the Oracle E-Business Suite Learning Management application can be crawled and indexed to readily search the courses offered, location and other details pertaining to the courses.

To crawl and index courses in Oracle E-Business Suite Learning Management:

1. Generate XML feed containing the courses. Each course can be an item in the feed. The properties of the course such as location and instructor can be set as attributes of the item.

2. Move the feed to a location accessible to Oracle SES through HTTP, FTP, or file protocol.

3. Generate a control file that points to that feed.

4. Generate a configuration file to point to this feed. Specify the feed type as control, the URL of the control feed, and the source name in the configuration file.

5. Create an **Oracle E-Business Suite 12** source in Oracle SES, specifying in the parameters the location of the configuration file, the user ID and the password to access the feed.

XML Configuration File

The configuration file is an XML file conforming to a set schema.

The following is an example of a configuration file to set up an XML-based source:

```
<rsscrawler xmlns="http://xmlns.oracle.com/search/rsscrawlerconfig">
    <feedLocation>ftp://my.host.com/rss_feeds</feedLocation>
    <feedType>directoryFeed</feedType>
    <errorFileLocation>/tmp/errors</errorFileLocation>
    <securityType>attributeBased</securityType>
    <sourceName>Contacts</sourceName>
    <securityAttribute name="EMPLOYEE_ID" grant="true"/>
</rsscrawler>
```

Where

- `feedLocation` is one of the following:

  - URL of the directory, if the data feed is a directory feed

    This URL should be the FTP URL or the file URL of the directory where the data feeds are located. For example:

    ```
    ftp://example.domain.com/relativePathOfDirectory
    file://example.domain.com/c:\dir1\dir2\dir3
    file://example.domain.com//private/home/dir1/dir2/dir3
    ```

    File URL if the data feeds are available on the same computer as Oracle SES. The path specified in the URL should be the absolute path of the directory.

FTP URL to access data feeds on any other computer. The path of the directory in the URL can be absolute or relative. The absolute path should be specified following the slash (/) after the host name in the URL. The relative path should be specified relative to the home directory of the user used to access FTP feeds.

The user ID used to crawl the source should have write permissions on the directory, so that the data feeds can be deleted after crawl.

- URL of the control file, if the data feed is a control feed

  This URL can be HTTP, HTTPS, file, or FTP URL. For example:

  ```
  http://example.com:7777/context/control.xml
  ```

  The path in FTP and file protocols can be absolute or relative.

- `feedType` indicates the type of feed. Valid values are `directoryFeed`, `controlFeed`, and `dataFeed`.

- `errorFileLocation` (optional) specifies the directory where status feeds should be uploaded.

  A status feed is generated to indicate the status of the processing feed. This status feed is named *data_feed_file_name*.suc or *data_feed_file_name*.err depending on whether the processing was successful. Any errors encountered are listed in the error status feed. If a value is specified for this parameter, then the status feed is uploaded to this location. Otherwise, the status feed is uploaded to the same location as the data feed.

  The user ID used to access the data feed should have write permission on the directory.

  If `feedLocation` is an HTTP URL, then `errorFileLocation` also should be an HTTP URL, to which the status feeds are posted. If no value is specified for `errorFileLocation`, then the status feeds are posted to the URL given in `feedLocation`.

  If an error occurs while processing a feed available over file or FTP protocol, then the erroneous feed is renamed *filename*.prcsdErr in the same directory.

- `sourceName` (optional) specifies the name of the source.

- `securityType` (optional) specifies the security type. Valid values are the following:

  - `noSecurity`: There is no security information associated with this source at the document level. This is the default value.

  - `identityBased`: Identity-based security is used for documents in the feed.

  - `attributeBased`: Attribute-based security is used for documents in the feed. With this security model, security attributes should be specified in the `securityAttribute` tag, and the values for these attributes should be specified for each document.

- `securityAttribute` specifies attribute-based security. One or more tags of this type should be specified, and each tag should contain the following attributes:

  - `name`: Name of the security attribute.

  - `grant`: Boolean parameter indicating whether this is a grant/deny attribute. The security attribute is considered a grant attribute if the value is true and a deny attribute if the value is false.

> **See Also:** "Configuration File XML Schema Definition" on page A-1

Overview of Document Service

A *document service* accepts input from documents and performs some operations on it, for example, you can create a document service for auditing. A document service can be used to perform the following operations for a document that needs to be indexed:

- Specifies to crawler whether to index a document or not.

- Adds, deletes, and updates document attributes and its values. The title, author, and description attribute values are always used for displaying search results.

- Replaces a document's original content with a revised content.

- Sets a document's language.

When you create a document service, you select the type of *document service manager*. You can either create a new document service manager or select from the list of existing document service managers defined in Oracle SES. The following document service managers are predefined in Oracle SES:

- **Secure Enterprise Search Document Summarizer:** This document service manager is used by default for search result clustering. It extracts the most significant phrases, and optionally sentences, from a document.

  > **See Also:** "Configuring and Using Clustering in Search Results" on page 5-7

- **Secure Enterprise Search Image Document Service:** This document service manager is used to extract metadata from image files, such as, XMP, EXIF, IPTC, and DICOM.

- **WebCenter Document Service:** This document service manager is used to process WebCenter documents.

A document service manager is defined using a *document service plug-in*, which is a Java class that implements the document service API. A document service plug-in accepts documents' content and attributes, and comes up with a revised document content and attributes.

  > **See Also:** "Document Service API" on page 13-44 for more information about creating a document service plug-in Java class

When a document service is invoked, the document parsing, attribute extraction, and language detection has been done. The crawler only honors the change made by the document service plug-in, and then the document is cached for indexing.

You can create multiple instances of a document service plug-in, that is, multiple document service instances. All the document service instances of a plug-in have the same configurable parameters, but may have different parameter values to cater to the specific requirements.

A *document service pipeline* is a collection of document service instances that are invoked in a particular order. The same instance can be assigned to different pipelines, but it cannot be assigned twice in the same pipeline. You can have multiple pipeline definitions; for example, one could be used globally and the other used for certain sources. Not every instance must be in a pipeline. You must perform a force re-crawl on a source if you add or change the document service pipeline for that source.

You can create document service managers, instances and pipelines using the Administration GUI page **Global Settings - Document Services**. You can set a global pipeline for all sources on the **Global Settings - Crawler Configuration** page. You can configure individual sources to use a particular pipeline on the **Home - Sources - Crawling Parameters** page. If enabled, the global pipeline is used for all the sources, unless a local service pipeline is defined.

You can also use the Administration API objects - `docServiceManager`, `docServiceInstance, and docServicePipeline` - for creating document service managers, instances and pipelines.

Creating a Document Service

The following are the steps to create a document service:

1. **Create a document service plug-in Java Class:**

   You must first create a document service plug-in Java class and store the implemented class in a jar file. Refer to the section "Document Service API" on page 13-44 for more information about creating a document service plug-in Java class.

2. **Create a document service instance:**

   The following are the steps to create an instance for a document service:

   a. In the Administration GUI, click **Global Settings - Document Services**.

   b. On the Document Service page, click **Create** button under the **Document Services** section.

   c. On the Create Document Service page, either select an already defined documentation manager from the **Available Managers** list or create a new documentation manager by selecting the **Create New Manager** option.

   d. Click **Next**.

   e. On the Create Document Service Manager page, enter the document manager class name and its jar file path of the document service plug-in Java class that was created in step 1.

   Note: Enter the full name of the document manager class including the package path, example, `oracle.search.plugin.doc.extractor.MyDocumentManager`.

   If the document manager jar file is placed in the *ses_home*`/search/lib/plugins/doc` directory, then its relative file path can be specified, else its absolute file path must be specified.

   ---

   f. Click **Next**.

   g. On the Create Document Service Instance page, enter the name for the document service instance and enter the appropriate values for its parameters, which control the way data needs to be extracted from text documents and image files.

   h. Click **Create**.

3. **Specify the order for invoking document service instances using a pipeline:**

The following are the steps to specify the order for invoking document service instances using a pipeline:

a. In the **Document Services Pipelines** section of the Document Service page, either click the Edit icon to update an existing pipeline or click the **Create** button to create a new pipeline.

b. In the Create/Edit Document Service Pipeline page, enter the name and description for the pipeline. In the Document Services section, select the document service instances from the list of already defined instances to build the pipeline. The document service instances will be invoked in the order in which they are listed in the pipeline.

c. Click **Create** for creating a new pipeline. Click **Apply** for updating an existing pipeline.

Configuring Support for Image Metadata

The Oracle SES crawler initially is set to search only text files. You can change this behavior by configuring an image document service connector to search the metadata associated with image files. Image files can contain rich metadata that provide additional information about the image itself.

The Image Document Service connector integrates Oracle Multimedia (formerly Oracle *inter*Media) technology with Oracle SES, for example, it enables extraction of metadata from image file, such as the medical X-ray image in DICOM format. This connector is separate from any specific data source.

The following table identifies the metadata formats (EXIF, IPTC, XMP, DICOM) that can be extracted from each supported image format (JPEG, TIFF, GIF, JPEG 2000, DICOM).

	JPEG	TIFF	GIF	JPEG2000	DICOM
EXIF	Yes	Yes	No	No	No
IPTC	Yes	Yes	No	No	No
XMP	Yes	Yes	Yes	Yes	No
DICOM	No	No	No	No	Yes

> **See Also:** *Oracle Multimedia User's Guide* and *Oracle Multimedia Reference* for more information about image metadata

Identifying Search Attributes for Image Metadata

Image files can contain metadata in multiple formats, but not all of it is useful when performing searches. A configuration file in Oracle SES enables you to control the metadata that is searched and published to an Oracle SES Web application.

The default configuration file is named `attr-config.xml`. Note that if you upgraded from a previous release, then the default configuration file remains `ordesima-sample.xml`.

You can either modify the default configuration file or create your own file. The configuration file must be located at *ses_home*`/search/lib/plugins/doc/ordim/config/`. Oracle recommends that you create a copy of the default configuration file before editing it. Note that the configuration file

must conform to the XML schema *ses_*
home/search/lib/plugins/doc/ordim/xsd/ordesima.xsd.

Oracle SES indexes and searches only those image metadata tags that are defined within the metadata element (between <metadata>...</metadata>) in the configuration file. By default, the configuration file contains a set of the most commonly searched metadata tags for each of the file formats. You can add other metatags to the file based on your specific requirements.

Image files can contain metadata in multiple formats. For example, an image can contain metadata in the EXIF, XMP, and IPTC formats. An exception to this are DICOM images, which contain only DICOM metadata. Note that for IPTC and EXIF formats, Oracle Multimedia defines its own image metadata schemas. The metadata defined in the configuration file must conform to the Oracle Multimedia defined schemas.

Because different metadata formats use different tags to refer to the same attribute, it is necessary to map metatags and the search attributes they define. Table 4–1 lists some commonly used metatags and how they are mapped in Oracle SES.

Table 4–1   Metatag Mapping

Oracle SES Attribute Name	Oracle SES Predefined Name	EXIF Metatag	IPTC Metatag	XMP Metatag
Author	Author	Artist	Author	photoshop:Creator
AuthorTitle	X	X	AuthorTitle	photoshop:AuthorsPosition
Description	Description	ImageDescription	Caption	dc:Description
Title	Title	X	ObjectName	dc:Title
Description Writer	X	X	captionWriter	photoshop:CaptionWriter
Headline1	Headline1	X	Headline	photoshop:Headline
Category	X	X	Category	photoshop:Category
Scene	X	X	X	Iptc4xmpCore:Scene
Publisher	X	X	X	dc:Publisher
Source	X	X	Source	photoshop:Source
Copyright	X	Copyright	Copyright	dc:rights
Keywords	Keywords	X	Keyword	dc:subject
Provider	X	X	Credit	photoshop:Credit
City	X	X	City	photoshop:City
State	X	X	provinceState	photoshop:State
Country	X	X	Country	photoshop:Country
Location	X	X	location	Iptc4xmpCore:Location
EquipmentMake	X	Make	X	tiff:Make
EquipmentModel	X	Model	X	tiff:Model

Oracle SES provides this mapping in the configuration file `attr-config.xml`. You can edit the file to add other metatags. Oracle recommends that you make a copy of the original configuration file before editing the settings. The configuration file defines the display name of a metatag and how it is mapped to the corresponding metadata in each of the supported formats.

This is done using the `searchAttribute` tag, as shown in the example below:

```
<searchAttribute>
 <displayName>Author</displayName>
 <metadata>
   <value format="iptc">byline/author</value>
   <value format="exif">TiffIfd/Artist</value>
   <value format="xmp">dc:creator</value>
   <value format="xmp">tiff:Artist</value>
 </metadata>
</searchAttribute>
```

For each search attribute, the value of `displayName` is an Oracle SES attribute name that is displayed in the Oracle SES web application when an Advanced Search - Attribute Selection is performed. If any of the listed attributes are detected during a crawl, then Oracle SES automatically publishes the attributes to the SES web application.

For the element `value`, `format` must take the value of a supported format such as `iptc`, `exif`, `xmp`, or `dicom`.

The value defined within the element, for example, `byline/author`, is the XML path when the image format is IPTC, EXIF, or XMP. For DICOM, this value must be the standard tag number or value locator.

For IPTC and EXIF formats, the XML path must conform to the metadata schemas defined by Oracle Multimedia. These schemas are defined in the files `ordexif.xsd` and `ordiptc.xsd` located at *ses_home*/search/lib/plugins/doc/ordim/xsd/.

You do not need to specify the root elements defined in these schemas (`iptcMetadata`, `exifMetadata`) in the configuration file. For example, you can specify `byline/author` as the `xmlPath` value of the author attribute in IPTC format. Oracle Multimedia does not define XML schemas for XMP metadata, so refer to the Adobe XMP specification for the `xmlPath` value.

Within the `<searchAttribute>` tag, you can also specify an optional `<dataType>` tag if the attribute carries a date or numeric value. For example,

```
<searchAttribute>
     <displayName>AnDateAttribute</displayName>
     <dataType>date</dataType>
     <metadata>
        ...
     </metadata>
</searchAttribute>
```

The default data type is string, so you do not have to explicitly specify a string.

Supporting XMP Metadata

Oracle SES supports both standard and custom XMP metadata searches. Because all XMP properties share the same parent elements `<rdf:rdf><rdf:description>`, you must specify only the real property schema and property name in the configuration file. For example, specify `photoshop:category` instead of `rdf:rdf/rdf:description/photoshop:category`. The same rule applies to XMP

custom metadata also. However, for XMP structure data, you must specify the structure element in the format parent/child 1/child 2/…child N, where child N is a leaf node. For example, `Iptc4xmpCore:CreatorContactInfo/Iptc4xmpCore:CiPerson`. Note that the image plug-in does not validate the metadata value for XMP metadata.

XMP metatags consist of 2 components separated by a colon(:). For example, `photoshop:Creator` corresponds to the `Author` attribute (see Table 4–1). In this example, `photoshop` refers to the XMP schema namespace. The other common namespaces include `dc`, `tiff`, and `Iptc4xmpCore`.

Before defining any XMP metadata in the configuration file, you must ensure that the namespace is defined. For example, before defining the metadata `photoshop:Creator`, you must include the namespace `photoshop` in the configuration file. This rule applies to both the standard and custom XMP metadata namespaces. As a best practice, Oracle recommends that you define all the namespaces at the beginning of the configuration file. If the namespace defined in the configuration file is different from the one in the image, then Oracle SES cannot find the attributes associated with this namespace. You can define namespaces as shown:

```
<xmpNamespaces>
<namespace
prefix="Iptc4xmpCore">http://iptc.org/std/Iptc4xmpCore/1.0/xmlns/</namespace>
<namespace prefix="dc">http://purl.org/dc/elements/1.1/</namespace>
<namespace prefix="photoshop">http://ns.adobe.com/photoshop/1.0/</namespace>
<namespace prefix="xmpRights">http://ns.adobe.com/xap/1.0/rights/</namespace>
<namespace prefix="tiff">http://ns.adobe.com/tiff/1.0/</namespace>
</xmpNamespaces>
```

Note that the Adobe XMP Specification requires that XMP namespaces end with a slash (/) or hash (#) character.

> **See Also:** *Adobe Extensible Metadata Platform (XMP) Specification* for the XMP metadata schema and a list of standard XMP namespace values:
>
> http://www.adobe.com/devnet/xmp.html

Custom XMP metadata must be explicitly added to `attr-config.xml`. An example of a custom metadata is:

```
<xmpNamespaces>
  <namespace prefix="hm">http://www.oracle.com/ordim/hm/</namespace>
</xmpNamespaces>
<searchattribute>
  <displayname>CardTitle</displayname>
  <metadata>
    <value format="xmp">hm:cardtitle</value>
  </metadata>
</searchattribute>
```

Supporting DICOM Metatags

Oracle SES 11*g* supports DICOM metatags, and these metatags are available in the default configuration file `attr-config.xml`. Note that the configuration file `ordesima-sample.xml`, which is the default configuration file if you upgraded from a previous release, does not contain DICOM metatags. Therefore, you must manually add DICOM metatags to the `ordesima-sample.xml` file. You can copy the DICOM metatags from `attr-config.xml`, which is available in the same directory. You can also reference the DICOM standard and add additional DICOM tags.

DICOM metatags are either DICOM standard tags or DICOM value locators.

DICOM Standard Tags DICOM standard tags are 8-digit hexadecimal numbers, represented in the format `ggggeeee` where `gggg` specifies the group number and `eeee` specifies the element number. For example, the DICOM standard tag for the attribute `performing physician's name` is represented using the hexadecimal value 00081050.

Note that the group number `gggg` must take an Even value, excepting 0000, 0002, 0004, and 0006, which are reserved group numbers.

The DICOM standard defines over 2000 standard tags.

The file `attr-config.xml` contains a list of predefined DICOM standard metatags. You can add new metatags to the file as shown in the following example:

```
<searchAttribute>
      <displayName>PerformingPhysicianName</displayName>
      <metadata>
        <value format="dicom">00081050</value>
      </metadata>
</searchAttribute>
```

> **Note:** The image connector does not support SQ, UN, OW, OB, and OF data type tags. Therefore, do not define such tags in the configuration file.

> **See Also:** http://medical.nema.org for more information about the standard tags defined in DICOM images, and the rules for defining metatags

DICOM Value Locators Value locators identify an attribute in the DICOM content, either at the root level or from the root level down.

A value locator contains one or more sublocators and a tag field (optional). A typical value locator is of the format:

```
sublocator#tag_field
```

Or of the format:

```
sublocator
```

Each sublocator represents a level in the tree hierarchy. DICOM value locators can include multiple sublocators, depending on the level of the attribute in the DICOM hierarchy. Multiple sublocators are separated by the dot character (.). For example, value locators can be of the format:

```
sublocator1.sublocator2.sublocator3#tag_field
```

Or of the format:

```
sublocator1.sublocator2.sublocator3
```

A `tag_field` is an optional string that identifies a derived value within an attribute. A tag that contains this string must be the last tag of a DICOM value locator. The default is `NONE`.

A sublocator consists of a `tag` element and can contain other optional elements. These optional elements include `definer` and `item_num`. Thus, a sublocator can be of the format:

```
tag
```

Or it can be of the format

```
tag(definer)[item_num]
```

Table 4–2   Sub Components of a Sublocator

Component	Description
tag	A DICOM standard tag represented as an 8-digit hexadecimal number.
definer	A string that identifies the organization creating the tag. For tags that are defined by the DICOM standard, the default value (which can be omitted) is DICOM.
	Note that Oracle SES supports DICOM standard tags alone. It does not support private tags.
item_num	An integer that identifies a data element within an attribute, or a wildcard character ("*") that identifies all data elements within an attribute. It takes a default value of 1, the first data element of an attribute. This parameter is optional.

The following example shows how to add a value locator to the `attr-config.xml` file:

```
<searchAttribute>
  <displayName>PatientFamilyName</displayName>
  <metadata>
  <value format="dicom">00100010#UnibyteFamily</value>
  </metadata>
</searchAttribute>
```

where `UnibyteFamily` is a tag_field of person name.

The following example shows how to define a value locator from the root level.

```
<searchAttribute>
     <displayName>AdmittingDiagnosisCode</displayName>
     <metadata>
       <value format="dicom">00081084.00080100</value>
     </metadata>
</searchAttribute>
<searchAttribute>
     <displayName>AdmittingDiagnosis</displayName>
     <metadata>
       <value format="dicom">00081084.00080104</value>
     </metadata>
</searchAttribute>
```

In the above example, the tag 00081084 represents the root tag `Admitting Diagnoses Code Sequence`. This tag includes four child tags: `code value` (0008, 0100), `coding scheme designator` (0008, 0102), `coding scheme version` (0008, 0103) and `code meaning` (0008, 0104). This example defines the value locators for `code value`: 00081084.00080100 and `code meaning`: 00081084.00080104.

> **Note:** The image connector does not support SQ, UN, OW, OB, and OF data type value locators. Therefore, ensure that the last sublocator of a value locator does not specify such data types.

> **See Also:** *Oracle Multimedia DICOM Developer's Guide* for more information about DICOM value locators

Example: Adding an Attribute to the Default attr-config.xml File

To search for information about image caption writer:

1. Open Oracle SES Administration GUI and create the DescriptionWriter attribute:

   Specify **DescriptionWriter** as an Oracle SES attribute name (shown on the Advanced Search - Attribute Selection page).

2. Examine the following sources for information relevant to modifying the default `attr-config.xml` file:

   - Oracle Multimedia IPTC schema at *ses_home*/search/lib/plugins/doc/ordim/xsd/ordiptc.xsd. The IPTC metadata for image caption writer is shown as `captionWriter`.

   - Adobe XMP Specification for XMP Metadata. The XMP path for this property is defined as `photoshop:CaptionWriter`.

   - Oracle Multimedia EXIF schema. There is no caption writer metadata in EXIF.

3. Add the following section to `attr-config.xml`:

```
<searchAttribute>
    <displayName>DescriptionWriter</displayName>
    <metadata>
        <xmlPath format="iptc">captionWriter</xmlPath>
        <xmlPath format="xmp">photoshop:CaptionWriter</xmlPath>
    </metadata>
</searchAttribute>
```

4. If the `photoshop` XMP namespace is not registered in the configuration file, then add the `namespace` element to `xmpNamespaces` as shown here:

```
<xmpNamespaces>
    <namespace prefix="photoshop">http://ns.adobe.com/photoshop/1.0/</namespace>
        .
        .   existing namespaces
        .
</xmpNamespaces>
```

Creating Image Document Service Connector

A default Image Document Service connector instance is created during the installation of Oracle SES. You can configure the default connector or create a new one.

To create an Image Document Service instance:

1. In the Oracle SES Administration GUI, click **Global Settings**.

2. Under Sources, click **Document Services** to display the Global Settings - Document Services page.

3. To configure the default image service instance:

   a. Click **Expand All**

   b. Click **Edit** for the default image service instance.

   or

   To create a new image service instance:

   a. Click **Create** to display the Create Document Service page.

   b. For **Select From Available Managers**, choose **ImageDocumentService**. Provide a name for the instance.

4. Provide a value for the **attributes configuration file** parameter.

   The default value of **attributes configuration file** is `attr-config.xml`. The file is located at *ses_home*`/search/lib/plugins/doc/ordim/config/`, where *ses_home* refers to the directory which stores the Oracle SES specific components. If you create a new configuration file, then you must place the file at the same default location.

5. Click **Apply**.

6. Click **Document Services** in the locator links to return to the Document Services page.

7. Add the Image Document Service plug-in to either the default pipeline or a new pipeline.

To add the default Image Document Service plug-in to the default pipeline:

1. Under Document Service Pipelines, click **Edit** for the default pipeline.

2. Move the Image Document Service instance from **Available Services** to **Used in Pipeline**.

3. Click **Apply**.

To create a new pipeline for the default Image Document Service plug-in:

1. Under Document Service Pipelines, click **Create** to display the Create Document Service Pipeline page.

2. Enter a name and description for the pipeline.

3. Move the Image Document Service instance from **Available Services** to **Used in Pipeline**.

4. Click **Create**.

Using Image Document Service Connector

You must either create a source to use the connector or enable the connector for an existing source.

To enable the connector for an existing source:

1. Click **Sources** on the home page.

2. Click the **Edit** icon for the desired source.

3. Click **Crawling Parameters**.

4. Select the pipeline that uses the Image Document Service and enable the pipeline for this source.

5. Click **Document Types**. From the **Not Processed** column, select the image types to search and move them to the **Processed** column. The following sources are supported: JPEG, JPEG2000, GIF, TIFF, DICOM.

Searching Image Metadata

You can search image metadata from either the Oracle SES Basic Search page or the Advanced Search - Attribute Selection page.

For Basic Search, Oracle SES searches all the metadata defined in the configuration file for each supported image document (JPEG, TIFF, GIF, JPEG 2000, and DICOM). It returns the image document if any matching metadata is found.

Advanced Search enables you to search one or more specified attributes. It also supports basic operations for date and number attributes. Oracle SES returns only those image documents that contain the specified metadata.

Note that Oracle SES does not display the Cache link for image search results.

Troubleshooting Image Document Service Connector

If the Image Document Service Connector fails, then check the following:

- Is the pipeline with an Image Document Service connector instance enabled for the source?

- Are the image types added to the source?

- For a web source, are the correct MIME types included in the HTTP server configuration file?

  For example, if you use Oracle Application Server, then check the file *ses_home*/`Apache/Apache/conf/mime.types`. If the following media types are missing, then add them:

MIME Type	Extensions
image/jp2	jp2
application/dicom	dcm

- If a connection is established, and not all the image files are crawled, then check whether the recrawl policy is set to `Process Documents That Have Changed`. If so, change the policy to `Process All Documents`:

  1. Go to Home - Schedules.

  2. Under Crawler Schedules, click **Edit** for the specific source to open the Edit Schedule page.

  3. Under Update Crawler Recrawl Policy, select **Process All Documents**.

  You can change the recrawl policy back to **Process Documents That Have Changed**, after the crawler has finished crawling all the documents in the new source.

Overview of Attributes

Each source has its own set of document attributes. Document attributes, like metadata, describe the properties of a document. The crawler retrieves values and maps them to search attributes. This mapping lets users search documents based on

their attributes. Document attributes in different sources can be mapped to the same search attribute. Therefore, users can search documents from multiple sources based on the same search attribute.

After you crawl a source, you can see the attributes for that source. Document attribute information is obtained differently depending on the source type.

Document attributes can be used in tasks such as document management, access control, or version control. Different sources can have different attribute names that are used for the same idea; for example, version and revision. It can also have the same attribute name for different ideas; for example, *language* as in natural language in one source but as programming language in another.

You can view all the document attributes defined in Oracle SES, along with their data sources and search attributes, in the **Global Settings - Document Attributes** page of Administration GUI.

Oracle SES has several default search attributes. They can be incorporated in the query application for a more detailed search and richer presentation.

Search attributes are defined in the following ways:

- System-defined search attributes, such as title, author, description, subject, and mimetype.

- Search attributes created by the Oracle SES administrator.

    > **Note:** For a Web data source with a custom metatag that needs to be treated as a search attribute, the administrator must define the search attribute first and then map the metatag to the search attribute for that data source in the **Global Settings - Document Attributes** page of the Administration GUI. A metatag can only be mapped to a String type search attribute. Oracle SES does not support mapping of Date type and Number type attributes.

- Search attributes created by the crawler. During crawling, the crawler plug-in maps document attributes to search attributes that have the same name and data type. If not found, then the crawler creates a new search attribute with the same name and type as the document attribute defined in the crawler plug-in, only for the default metatags like title and author.

    > **Note:** Search attribute names must be unique; two attributes cannot have the same name. For example, if a search attribute exists with a String data type, and another search attribute is discovered by the crawler with the same name but a different data type, then the crawler ignores the second attribute.
    >
    > To prevent this conflict and allow Oracle SES to index both attributes, check the list of Oracle SES attribute names and types in Administration GUI before creating new attributes.

Attributes For Different Source Types

Table and database sources have no predefined attributes. The crawler collects attributes from columns defined during source creation. You must map the columns to the search attributes.

For Siebel 7.8 sources, specify the attributes in the query while creating the source. For Oracle E-Business Suite and Siebel 8 sources, specify the attributes in the XML data feed.

For many source types, such as OracleAS Portal, e-mail, NTFS, and Microsoft Exchange sources, the crawler picks up key attributes offered by the target systems. For other sources, such as Lotus Notes, an **Attribute list** parameter is in the Home - Sources - Customize User-Defined Source page. Any attributes that you define are collected by the crawler and available for search.

Using Lists of Values for Search Attributes

The list of values (LOV) for a search attribute can help you specify a search. Global search attributes can be specified on the Global Settings - Search Attributes page. For user-defined sources where LOV information is supplied through a crawler plug-in, the crawler registers the LOV definition. Use the Oracle SES Administration GUI or the crawler plug-in to specify attribute LOVs, attribute value, attribute value display name, and its translation.

When multiple sources define the LOV for a common attribute, such as title, the user sees all the possible values for the attribute. When the user restricts search within a particular source group, only LOVs provided by the corresponding sources in the source group are shown.

LOVs can be collected automatically. The following example shows Oracle SES collecting LOV values to crawl a fictitious URL.

1. Create a Web source with `http://www.example.com` as the starting URL. Do not start crawling yet.

2. From the Global Settings - Search Attributes page, select the Attribute for Oracle SES to collect LOVs and click **Manage Lov**. (For example, click **Manage Lov** for **Author**.)

3. Select **Source-Specific** for the created source, and click **Apply**.

4. Click **Update Policy**.

5. Choose **Document Inspection** and click **Update**, then click **Finish**.

6. From the Home - Schedules page, start crawling the Web source. After crawling, the **LOV** button in the Advanced Search page shows the collected LOVs.

System-Defined Search Attributes

There are also two system-defined search attributes, `Urldepth` and `Infosource Path`.

`Urldepth` measures the number of levels down from the root directory. It is derived from the URL string. In general, the depth is the number of slashes, not counting the slash immediately following the host name or a trailing slash. An adjustment of -2 is made to home pages. An adjustment of +1 is made to dynamic pages, such as the example in Table 4–3 with the question mark in the URL.

`Urldepth` is used internally for calculating relevance ranking, because a URL with a smaller URL depth is typically more important.

Table 4–3 lists the `Urldepth` of some example URLs.

Table 4–3   Depth of Example URLs

URL	Urldepth
http://example.com/portal/page/myo/Employee_Portal/MyCompany	4
http://example.com/portal/page/myo/Employee_Portal/MyCompany/	4
http://example.com/portal/page/myo/Employee_Portal/MyCompany.htm	4
http://example.com/finance/finhome/topstories/wall_street.html?.v=46	4
http://example.com/portal/page/myo/Employee_Portal/home.htm	2

`Infosource Path` is a path representing the source of the document. This internal attribute is used in situations where documents can be browsed by their source. The `Infosource Path` is derived from the URL string.

For example, for this URL:

```
 http://example.com/portal/page/myo/Employee_Portal/home.htm
```

The `Infosource Path` is:

```
portal/page/myo/Employee_Portal
```

If the document is submitted through a connector, this value can be set explicitly by using the `DocumentMetadata.setSourceHierarchy` API.

Attribute Length Limitation for String Attributes

Oracle SES has a limitation of 4K bytes for custom string attribute values. When a custom string attribute value exceeds 4K bytes, Oracle SES truncates it to 4K bytes. Note that the attribute value length limit mentioned here is in terms of bytes and not in characters. Thus, for single-byte characters, the custom string attribute value length limit is 4K characters, while for multi-byte characters, this length limit is 1K characters.

Some default string attributes have smaller length limits, such as, `author` has 200 bytes, `description` has 300 bytes, `language` has 30 bytes, and security attributes have 1K bytes of length limit.

Understanding the Default Crawling Process

Oracle SES uses Pull type of crawling the sources by default. In the Pull type crawling, when the crawler runs for the first time, it must fetch data (Web pages, table rows, files, and so on) based on the source. It then adds the document to the Oracle SES index.

The Initial Crawl

This section describes a Web source crawling process for a schedule. It is divided into these phases:

- Queuing and Caching Documents
- Indexing Documents

Queuing and Caching Documents

The crawling cycle involves the following steps:

1. Oracle spawns the crawler according to the schedule you specify with the Oracle SES Administration GUI. When crawling is initiated for the first time, the URL queue is populated with the seed URLs.

2. The crawler initiates multiple crawling threads.

3. The crawler thread removes the next URL in the queue.

4. The crawler thread fetches the document from the Web. The document is usually an HTML file containing text and hypertext links. When the document is not in HTML format, the crawler converts the document into HTML before caching.

5. The crawler thread scans the HTML file for hypertext links and inserts new links into the URL queue. Duplicate links in the document table are discarded.

6. The crawler caches the HTML file.

7. The crawler registers the URL in the URL table.

8. The crawler thread starts over by repeating Step 3.

Fetching a document, as described in Step 4, can be time-consuming because of network traffic or slow Web sites. For maximum throughput, multiple threads fetch pages at any given time.

Indexing Documents

When the cache is full (default maximum size is 250 MB), the indexing process begins. At this point, the document content and any searchable attributes are pushed into the index.

When the **Preserve Document Cache** parameter is set to false, the crawler automatically deletes the cache after indexing the documents.

Note:

- Oracle SES indexes a document even if its file format is not recognized by the crawler. In this case only the document attributes are indexed for the connector crawl.

- Oracle SES does not index a document if its file size is more than the configured maximum allowed file size. There is a hard limit on this file size which is 1GB.

- Oracle SES always indexes built-in search attributes or default attributes that are defined in the Administration GUI page "Global Settings - Search Attributes".

- In RSS crawler, if a document and its attachment have the same attributes, and if the document does not contain any search attributes, then Oracle SES indexes the search attributes of the attachment.

- For attachments, Oracle SES by default does not index the attributes that are not defined as search attributes. To index such attributes, add these attributes as search attributes in the Administration GUI. Also, add these attributes in the Attribute Mapping tab for the data source that needs to be crawled and map these attributes to the search attributes created earlier.

Oracle SES Stoplist

Oracle SES maintains a stoplist. A stoplist is a list of words that are ignored during the indexing process. These words are known as stop words. stop words are not indexed because they are deemed not useful, or even disruptive, to the performance and accuracy of indexing. The Oracle SES stoplist contains only English words, and cannot be modified.

When you run a phrase search with a stop word in the middle, the stop word is not used as a match word, but it is used as a placeholder. For example, the word "on" is a stop word. If you search for the phrase "oracle on demand", then Oracle SES matches a document titled "oracle on demand" but not a document titled "oracle demand". If you search for the phrase "oracle on on demand", then Oracle SES matches a document titled "oracle technology on demand" but not a document titled "oracle demand" or "oracle on demand".

Maintenance Crawls

After the initial crawl, a URL page is only crawled and indexed if it changed since the last crawl. The crawler determines whether it has changed with the HTTP If-Modified-Since header field or with the checksum of the page. URLs that no longer exist are marked and removed from the index.

To update changed documents, the crawler uses an internal checksum to compare new Web pages with cached Web pages. Changed Web pages are cached and marked for re-indexing.

Data synchronization involves the following steps:

1. Oracle spawns the crawler according to the schedule specified in the Oracle SES Administration GUI. The URL queue is populated with the seed URLs of the source assigned to the schedule.

2. The crawler initiates multiple crawling threads.

3. Each crawler thread removes the next URL in the queue.

4. Each crawler thread fetches a document from the Web. The page is usually an HTML file containing text and hypertext links. When the document is not in HTML format, the crawler converts the document into HTML before caching.

5. Each crawler thread calculates a checksum for the newly retrieved page and compares it with the checksum of the cached page. If the checksum is the same, then the page is discarded and the crawler goes to Step 3. Otherwise, the crawler continues to the next step.

6. The crawler thread scans the document for hypertext links and inserts new links into the URL queue. Links that are in the document table are discarded. Oracle SES does not follow links from filtered binary documents.

7. The crawler marks the URL as accepted. The URL is crawled in future maintenance crawls.

8. The crawler registers the URL in the document table.

9. If the cache is full or if the URL queue is empty, then caching stops. Otherwise, the crawler thread starts over at Step 3.

A maintenance or a forced recrawl does not move a cache from the file system to the database, or the reverse. The cache location for a source remains the same until it is migrated to a different location.

Automatic Forced Recrawls

Changing the crawler configuration **Index Dynamic Page** from `No` to `Yes` for a Web source triggers an automatic forced recrawl of the data source.

The operation sets the Force Recall flag, but no notice is given of this change in mode.

Understanding the Push Type Crawling Process

In the Push type crawling, data sources push or publish the documents to Oracle SES server. The *pushed* documents are indexed almost in real time in Oracle SES and hence are immediately available for searching.

A data source that uses Push type crawling is called **Push Feed** data source. A **Push Feed** data source can be created in Oracle SES by either using the Administration GUI by selecting the Push Feed source type on the **Home - Sources** page or using the Administration API object type `sources`.

> **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for information about creating a Push Feed data source using the object type `sources` of Administration API.

Push Type Crawling Process

In the Push type crawling, Oracle SES administrator does not have to maintain crawler schedules or launch crawler schedules manually. Push Feed data sources must *push* the documents in the form of Data Feed XML to the Oracle SES Push endpoint URL `http://host:port/search/crawl/push`, where *host* and *port* are the Oracle SES server host and server port respectively. Schedule associated with a Push Feed data source is either processing a data feed or is waiting to accept a data feed from the data source at any point in time. Thus, a schedule associated with a Push Feed data source is always in the *Executing* state.

> **See Also:** Appendix A for information about the Data Feed XML schema definition.

A secure communication channel must exist between the Push Feed data sources and Oracle SES server. Therefore, an HTTP POST request containing a Data Feed XML from a Push Feed data source to Oracle SES Push endpoint URL must contain the following HTTP header parameters:

- X-TrustedEntity-Username – the trust entity user name specified in the Push Feed source configuration

- X-TrustedEntity-Password – the trust entity password specified in the Push Feed source configuration

- X-DataSouce – the Push Feed source name

- X-BatchID – the batch Id of the Push Feed data

The following Java code example generates the HTTP header information containing the above HTTP header parameters:

```
public void doPost(HttpServletRequest req, HttpServletResponse resp) throws
IOException, ServletException {
    ...
    req.setHeader("X-TrustedEntity-Username", user_name);
    req.setHeader("X-TrustedEntity-Password", password);
    req.setHeader("X-DataSource", push_feed_source_name);
```

```
        req.setHeader("X-BatchID", batch_id);
        ...
}
```

Once the *pushed* documents in the form of Data Feed XML are received by Oracle SES server, the documents are indexed almost in real time in Oracle SES and hence are immediately available for searching.

Push Feed Data Source Configuration Parameters

The following are the Push Feed data source parameters that can be configured in Oracle SES Administration GUI and Administration API:

- Trusted Entity Username and Password: The HTTP header parameters to be used as authentication credentials to accept documents.

- Username and Password for Fetching Attachments: The credentials for accessing any attachments that might need to be indexed. These can be empty, if there are no attachments to be indexed.

- Error Log HTTP URL: The URL to which the error log is posted.

- Attachment Realm and Attachment Auth Type: The realm required for authentication, so as to access attachments, and the supported attachment authentication types. The supported attachment authentication types are `BASIC` and `DIGEST`.

- Security Attributes: Security attributes for providing access permissions to users and groups. Multiple attributes can be specified using comma separated values, for example, `attribute1, true, attribute2, false`.

- Scratch Directory: The temporary directory on the system used by Push crawler for storing the temporary status feeds, before posting them back to data sources.

Viewing Error Messages Generated by Push Type Crawling

The error messages generated by the Push crawler can be accessed using the HTTP GET request of the Oracle SES endpoint URL `https://host:port/search/crawl?error` accompanied with the above mentioned HTTP header parameters, where *host* and *port* are the Oracle SES server host and server port respectively.

> **See Also:** "Push Feed Data Source Error Message XML Schema" on page 4-31

Viewing Log Messages Generated by Push Type Crawling

The Push endpoint log messages are stored in the Oracle SES server logs. Refer to the section "Viewing Oracle SES Server Log Files" on page 12-24 for more information about Oracle SES server logs.

The Push crawler log messages are stored in the Oracle SES crawler logs. The crawler logs can be viewed on the **Home - Schedules - Crawler Schedules** page of the Administration GUI like any other data source. Refer to the section "Viewing Crawler Logs" on page 4-35 for more information about Oracle SES crawler logs.

Push Feed Data Source Error Message XML Schema

The following is the XML schema used by Push Feed type data source for error reporting:

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified" attributeFormDefault="unqualified">
   <xs:element name="Errors">
      <xs:complexType>
         <xs:sequence>
            <xs:element name="sourceName" type="xs:string"/>
            <xs:element name="batchId" type="xs:string"/>
            <xs:element name="Instance">
               <xs:complexType>
                  <xs:complexContent>
                     <xs:extension base="xs:anyType">
                        <xs:attribute name="name"/>
                     </xs:extension>
                  </xs:complexContent>
               </xs:complexType>
            </xs:element>
         </xs:sequence>
      </xs:complexType>
   </xs:element>
</xs:schema>
```

The following is an example of an error message generated by a Push Feed data source:

Example 4–1   Push Feed Data Source Error Message Example

```
<Errors>
   <sourceName>buscomp</sourceName>
   <batchId>2007-04-16_15-24-43</batchId>
   <Instance name="local">
      <rss xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" version="2.0">
         <channel>
            <title>Siebel Search Center</title>
            <link>http://www.oracle.com/apps/rss</link>
            <lastBuildDate>2007-04-16T15:24:43.000Z</lastBuildDate>
            <channelDesc xmlns="http://xmlns.oracle.com/orarss">
               <sourceName>buscomp</sourceName>
               <feedType>Success</feedType>
               <batchId>2007-04-16_15-24-43</batchId>
            </channelDesc>
         </channel>
      </rss>
   </Instance>
</Errors>
```

Configuring the Crawler Settings using Administration API

Most crawler configuration tasks are controlled in the Oracle SES Administration GUI, but certain features (like title fallback, character set detection, and indexing the title of multimedia files) are controlled only by the Administration API. Configuration of the crawler is described by the `crawlerSettings` object of the Administration API.

> **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for more information about the `crawlerSettings` object of the Administration API.

The crawler uses a set of codes to indicate the crawling result of the crawled URL. Besides the standard HTTP status codes, it also uses its own codes for non-HTTP related situations.

See Also: [Appendix B, "URL Crawler Status Codes"](#)

Configuring the Crawler Log

You can configure various crawler log settings using the following methods:

- [Configuring the Crawler Log Using Administration GUI](#)
- [Configuring the Crawler Log Using Administration API](#)
- [Configuring the Crawler Log Using ODL Script clexecutor.sh](#)

Configuring the Crawler Log Using Administration GUI

On the **Global Settings - Crawler Configuration** page, you can specify the level of detail in which you want the crawler to store the log messages in the log file, that is, whether to log everything or to log only the summary information. You can also specify the language in which the crawler must store the log messages.

Configuring the Crawler Log Using Administration API

You can use the following elements of the `crawlerSettings` object that is used inside a `source` object of Administration API for specifying the crawler log configurations for a particular source:

- `verboseLogging`: Controls the level of detail for logging messages. Specify `true` to record all the information. Specify `false` to record only the summary information.
- `logLanguage`: Specify the language in which the crawler must store the log messages.
- `logLevel`: Specify the log level for the crawler log. The valid log levels are `TRACE`, `DEBUG`, `INFO`, `WARN`, `ERROR`, and `FATAL`.

> **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for more information about the `crawlerSettings` object of Administration API.

Configuring the Crawler Log Using ODL Script clexecutor.sh

Oracle SES stores the crawler log configurations in the ODL script file *ses_home*/bin/clexecutor.sh. You can edit this file in a text editor to set the following crawler log configurations:

- `LOG_FORMAT`: The log format to use. The available values are: `ODL-XML`, `ODL10-XML`, and `ODL-Text`. Default is `ODL-Text`.
- `LOG_MAX_FILE_SIZE`: The maximum size in bytes for each log file. When a log file reaches this size, it is archived, and a new log file is created. Default is 104857600 bytes.
- `LOG_ROTATION_FREQUENCY`: The frequency, in minutes, for archiving the old log file and creating a new one. This value must be either a number in minutes, or one of the following values (case-insensitive): `hourly`, `daily`, or `weekly`. Default is `daily`.

> **Note:** If a value is specified for the `LOG_MAX_FILE_SIZE` parameter, then Oracle SES ignores the value specified for the `LOG_ROTATION_FREQUENCY` parameter.

- `LOG_LEVEL`: Sets the log level in the form of a number. ODL logs only those records with a level equal to or higher than the specified log level value. The meaningful name for each log level is as follows:

Log Level	Name	Description
0	TRACE	Trace messages
2	DEBUG	Debug messages
4	INFO (NOTIFICATION)	Informational messages (default)
6	WARN	Warning messages
8	ERROR (SEVERE)	Error messages
10	FATAL (INCIDENT_ERROR)	Fatal messages

- `LOG_AUTOFLUSH_LEVEL`: Sets the log level for auto-flushing. The ODL allows log records to be buffered till it encounters a log record with a level equal to or higher than the specified auto-flush level, and at that time, ODL automatically flushes the buffer. The available values are: `TRACE`, `DEBUG`, `NOTIFICATION`, `WARN`, `ERROR`, and `FATAL`. Default is `NOTIFICATION`.

> **Note:** Crawler log level can be changed using the `LOG_LEVEL` configuration setting specified in the `clexecutor.sh` file as well as using the `logLevel` element of the `crawlerSettings` object of the Administration API. The crawler uses the log level that is more granular of the two.

Monitoring the Crawling Process

You can monitor the crawling process by using the following methods:

- Check the crawl progress and crawl status on the **Home - Schedules** page (Click **Refresh Status**) of Administration GUI.

- Monitor your crawler statistics on the **Home - Statistics** page and the **Home - Schedules - Crawler Progress Summary** page of Administration GUI. See "Viewing Crawler Statistics" on page 4-34 for more information about viewing the crawler statistics.

- Monitor the crawler log file for the current schedule. See "Viewing Crawler Logs" on page 4-35 for more information about viewing the crawler log file.

> **See Also:** "Tuning Crawl Performance" on page 12-5

Viewing Crawler Statistics

The following crawler statistics are shown on the Home - Schedules - Crawler Progress Summary page. Some of them are also shown in the log file, under "Crawling results".

- **Documents Processed**: Number of documents retrieved by the crawler. Many of these documents may not have been queued for indexing yet.

- **Documents Discovered**: Total number of documents discovered so far. This is roughly equal to:

[documents processed] + [documents to process] +
[document processing failures] + [documents excluded]

In an RSS-based connector crawl like UCM, the number of documents or items from a data feed correspond to the number of documents discovered. This assumes that all of the data feed items are valid, that is, they have valid display URLs and so on.

- **Documents to Process**: Number of URLs in the queue waiting to be crawled.

- **Documents Deleted**: Number of document deleted during incremental recrawl.

- **Document Processing Failures**: Number of documents whose contents cannot be retrieved by the crawler. This could be due to an inability to connect to the Web site, slow server response time causing time-outs, or authorization requirements. Problems encountered after successfully fetching the document are not considered here; for example, documents that are too big or duplicate documents that were ignored.

- **Documents Excluded**: Number of URL links encountered but not considered for crawling. The reason for excluding these documents could be due to boundary rules, the robots exclusion rule, the mime type inclusion rule, the crawling depth limit, or the URL rewriter discard directive.

- **Documents Non-Indexable**: Number of documents that cannot be indexed; for example, a file source directory or a document with robots NOINDEX meta-tag.

- **Document Conversion Failures**: Number of binary documents that could not be converted to HTML. Though the filtering operation is failed for these documents, their metadata is submitted to the Oracle SES indexing engine (that is, Oracle Text) for indexing purpose. This document count is included in the "Documents Queued for Index" statistic.

- **Conversion Failures in Archives**: Number of archived/zipped documents with conversion failures. Conversion failure of each document inside a zip file is counted as one document failure.

- **Documents Queued for Index**: Number of documents that are pending to be indexed, that is, number of documents that are submitted to the Oracle SES indexing engine (that is, Oracle Text) for the indexing purpose and are yet to be indexed.

- **Documents Indexed**: Number of documents that have been indexed.

Viewing Crawler Logs

An Oracle SES log file records all the crawler activity, warnings, and error messages for a particular schedule. It includes messages logged at startup, run time, and shutdown.

> **Note:** Oracle SES 11.2.2.2 uses ODL as the default logger. The earlier Oracle SES release (11.1.2.2) used log4j as the default logger.

Viewing Logs using Administration GUI

You can view the crawler log on the **Home - Schedules - Crawler Schedules** page. Each logging message in the log is a one line message containing the tab delimited text having mainly the following fields. The number of fields appearing in the crawler log will vary depending upon the log level and the log format settings.

1. Timestamp

2. Message level

3. Crawler thread name

4. Component name and module name.

5. Message

Viewing Logs in a File System

A new log file is created whenever you restart the crawler. The location of the crawler log file can be found on the **Home - Schedules - Crawler Progress Summary** page in the Administration GUI. The crawler maintains the past seven versions of its log file. The most recent log file is shown in the Oracle SES Administration GUI. You can view all the log files in the file system.

The log file name format is:

```
search.crawler.i<SES_Instance_ID>ds<Data_Source_ID>.<timestamp>.log
```
where,

- *<SES_Instance_ID>* is the SID of the SES database.

- *<Data_Source_ID>* is the identifier of the data source being crawled.

- *<timestamp>* is the crawl starting time in Greenwich Mean Time (GMT) 24-hour format - MMDDHHmm, that is, month, day, hours, and minutes.

Viewing Logs using `viewlog` Command-line Tool

You can also view the crawler log files using the **viewlog** command-line tool available in the *ses_home*/bin directory. You can use the viewlog tool specially when the crawler log files are not accessible from the system where the Administration GUI application is installed.

The following is the syntax of viewlog command:

```
viewlog -log crawler_log_file -[options]
```
where,

crawler_log_file is the name of the crawler log file that you want to view. See "Viewing Logs in a File System" on page 4-36 for more information about the file name format for the crawler log.

[*options*] are the optional parameter-value pairs that you can specify in the viewlog command to filter and display specific information from the log file.

> **Note:** Run the viewlog -help command to know the detail syntax for using the viewlog tool.

The following list describes the various optional parameters supported by the viewlog command.

Optional Parameter	Description
help	Displays the syntax of the viewlog command along with the description of all the supported optional parameters.
type	Displays log messages of specific message types.
	Example:
	viewlog -log crawler_log_file -type ERROR WARNING
	This displays the log messages of only ERROR and WARNING types.

Optional Parameter	Description
pattern	Displays log messages matching a specific pattern. Example: `viewlog -log crawler_log_file -pattern "Exception*"` This displays the log messages that contain the word starting with `Exception` string.
last	Displays recent log messages logged-in within the specific time frame. Example: `viewlog -log crawler_log_file -last "2h"` This displays the log messages logged-in within the last 2 hours.
maxRecords	Displays the specific count of log messages on a page. `viewlog -log crawler_log_file -maxRecords 200` This displays 200 log messages on a page. Example:
groupBy	Displays log messages grouped by specific ODL message attributes. Example: `viewlog -log crawler_log_file -groupBy COMPONENT_ID MSG_TYPE` This displays the log messages grouped according to the `COMPONENT_ID` and `MSG_TYPE` ODL message attributes.
tail	Displays the specific count of recent log messages. Example: `viewlog -log crawler_log_file -tail 200` This displays the recent 100 log messages.
format	Displays log messages in a specifc format. The valid values are `ODL-Text`, `ODL-XML`, `ODL-complete` and `simple`. The default format is `ODL-Text`. Example: `viewlog -log crawler_log_file -format ODL-XML` This displays the log messages in XML format.
query	Displays log messages matching a specific boolean expression. Example: `viewlog -log crawler_log_file -query "MSG_TYPE eq ERROR or MSG_TYPE eq WARNING"` This displays only the `ERROR` and `WARNING` type of log messages.

Parallel Query Indexing

To scale up the indexed data size while maintaining satisfactory query response time, the indexed data can be stored in independent disks to perform disk I/O operations in parallel. The major features of this architecture are:

- Oracle SES index is partitioned, so that the sub-queries are executed in parallel.

- Disks perform I/O operations independent of one another. As a result, the I/O bus contention does not create a significant bottleneck on the collective I/O throughput.

■ Partition rules are used to control the document distribution among the partitions.

Figure 4–1 End User Query Partitioning
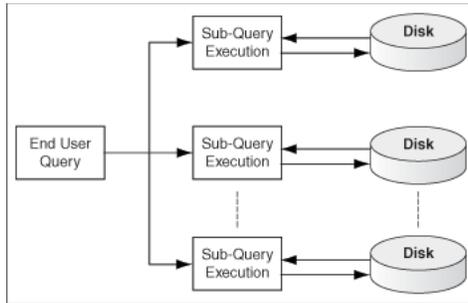
Document Partition Model and Storage Areas

A storage area is a database tablespaces in Oracle SES. Storage areas are used to store the partitions when the partitioning option is enabled in Oracle SES.

There are two kinds of partitioning mechanisms for improving query performance - attribute-based partitioning and hash-based partitioning.

Attribute-based partitioning uses attribute values to distribute documents into multiple partitions. For example, documents can be stored into different partitions based on their data source type, where data source type is used as an attribute for deciding partitions. Attribute-based partitioning is used for pruning purpose.

Hash-based partitioning uses a hash function to distribute a large set of documents into multiple partitions. For example, documents can be stored into different partitions based on the file path corresponding to each of the disks. Hash-based partitioning is used for executing parallel queries.

A partition engine controls the partition logic at both crawl time and query time. When a large data set must be searched without pruning the conditions, the end user request is broken into multiple parallel sub-queries so that the I/O and CPU resources can be utilized in parallel. After the result sets of the sub-queries are returned by the independent query processors, a merged result set is returned to the end user.

Figure 4–2 shows how the mechanism works during crawl time. The documents are partitioned and stored in different storage areas. Note that the storage areas are created on separate physical disks, so that I/O operations can be performed in parallel to improve the search turn around time.

Figure 4–2 Document Partitioning at Crawl Time
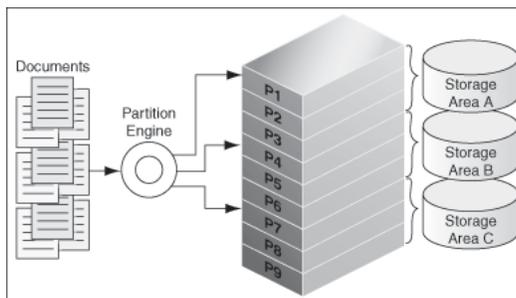
At query time, the query partition engine generates sub-queries and submits them to the storage areas, as shown in Figure 4–3.

Figure 4–3   Generation of Sub Queries at Query Time
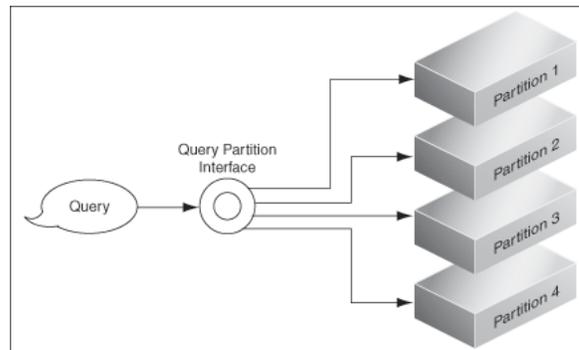
See "Parallel Querying and Index Partitioning" on page 12-15 for more information.

5

Customizing the Search Results

This chapter explains the various ways available for customizing the search results. It contains the following topics:

- Adding Suggested Content in Search Results
- Customizing the Appearance of Search Results
- Configuring and Using Clustering in Search Results
- Configuring and Using Facets in Search Results
- Configuring and Using Tags in Search Results
- Configuring Default Sort Conditions
- Configuring Top-N Documents and Group/Sort Attributes
- Configuring Sort Criteria using Sortable Attributes (Absolute Sort)
- Configuring and Using Source Groups in Search Results
- Customizing the Relevancy of Search Results
- Using Relevancy Boosting for URLs and Query Terms

Adding Suggested Content in Search Results

Suggested content lets you display real-time data content along with the result list in the default query application. Oracle SES retrieves data from content providers and applies a style sheet to the data to generate an HTML fragment. The HTML fragment is displayed in the result list and is available through the Web Services API. For example, when an end user searches for contact information on a coworker, Oracle SES can fetch the content from the suggested content provider and return the contact information (e-mail address, phone number, and so on) for that person with the result list. Suggested content results appear in tabbed panes above the query results. When the query returns no results, suggested content is not displayed.

Configure suggested content on the Search - Suggested Content page in the Oracle SES Administration GUI. Enter the maximum number of suggested content results (up to 20) to be included with the Oracle SES result list. The results are rendered on a first-come, first-served basis.

Suggested Content Providers

Regular expressions (as supported in the Java regular expression API `java.util.regex`) are used to define query patterns for suggested content providers. The regular expression-based pattern matching is case-sensitive. For example, a

provider with the pattern `dir\s(\S+)` is triggered on the query `dir james` but not on the query `Dir James`. To trigger on the query `Dir James`, the pattern could be defined either as `[Dd][Ii][Rr]\s+(\S+)` or as `(?i)dir\s+(\S+)`. A provider with a blank query pattern is triggered on all queries.

The URL you enter for the suggested content provider can contain the following variables: `$ora:q`, `$ora:lang`, `$ora:q1` ... `$ora:q`*n* and `$ora:username`.

- `$ora:q` is the end user full query.

- `$ora:lang` is the two-letter code for the browser language.

- `$ora:q`*n* is the *n*th regular expression match group from the end user query. *n* starts from 1. If no *n*th group is matched, then the empty string replaces the variable.

- `$ora:username` is the end user name.

Enter an XSLT style sheet to define rules (for example, the size and style) for transforming XML content from a provider into an HTML fragment. This HTML fragment is displayed in the result list or returned over the Web Services API. If you do not enter an XSLT style sheet, then Oracle SES assumes that the suggested content provider returns HTML. If you do not enter an XSLT style sheet and the provider returns XML, then the result list displays the plain XML.

> **Note:** As an administrator, you are responsible for verifying that the suggested content providers return valid and safe content. Corrupted or incomplete content returned by a suggested content provider can affect the formatting of the default query application results page.

Security Options

There are three security options for how Oracle SES passes the end user's authentication information to the suggested content provider:

- **None**: No security policy is used. (Default)

- **Cookie**: The end user first must be authenticated by the suggested content provider. A cookie is set for the user to maintain a session. Oracle SES must know the cookie used by the provider for authentication, and it is made available during registration of the suggested content provider. When the user enters a query, Oracle SES grabs the cookies from the user's request header and passes them to the provider. The cookie scope must be set to the common domain of the provider site and the Oracle SES site by the provider.

   For example, suppose the provider site is `http://provider.example.com` and the Oracle SES site is `http://ses.example.com`. After the end user logs in to the provider site, the site could set the value of the security cookie loginCookie with domain scope `.example.com`. When the end user searches in Oracle SES, Oracle SES gets the loginCookie value from the end user browser and forwards it to the provider site to get the suggested content (without login to the provider site again). However, if the provider site is accessed as `http://provider` or if the Oracle SES site is accessed as `http://SES`, then no domain cookie is available for sharing between the two sites and this security mechanism does not work.

   You can decide what happens when suggested content is available but the user is not logged in to the suggested content provider or the cookie for the provider is not available. For **Unauthenticated User Action**, if you select **Ignore content**, then content from that provider is not displayed in the result list. If you select **Display**

login message, then Oracle SES returns a message that there is content available from this provider but the user is not logged in. The message also provides a link to log in to that provider. Enter the link for the suggested content provider login in the **Login URL** field.

■ **Service-to-Service**: A one-way trusted relationship is established between Oracle SES and the suggested content provider. Any user logged in to Oracle SES does not need to be authenticated by the provider again. The provider only authenticates the Oracle SES application and trusts the Oracle SES application to act as the end user.

The end user identity is sent from Oracle SES to the provider site in the HTTP header `ORA_S2S_PROXY_USER`. The trusted entity could be a proxy user configured in the identity management system used by the provider, or it could be a name-value pair.

If the secured content provider must authenticate the end user and it sets the domain level security cookie to maintain login information after the end user login, then use the cookie method for form authentication. The Oracle SES end user must login manually to the provider site, and the security cookie is stored in the browser. Oracle SES searches on the provider for the end user without additional login.

However, if the domain security cookie is not allowed for the provider, then the provider must support service-to-service security. The provider must allow an Oracle SES application account to search after passing HTTP basic or digest authentication. Also, if the provider has different secured content for different Oracle SES end users, then it must respect the end user security (in the HTTP header `ORA_S2S_PROXY_USER`) for the Oracle SES search request.

To register a provider that requires either HTTP basic or HTTP digest authentication, specify the authentication user name in the **Entity Name** field and specify the authentication password in the **Entity Password** field.

Example Configuring Google OneBox for Suggested Content

Existing OneBox providers can be configured as Oracle SES suggested content providers. For example, for a Google OneBox provider, the provider URL might be `http://host.company.com/apps/directory.jsp` and the trigger might be `dir\s(\S+)`. When the user query is "dir james", the provider receives the request with a query string similar to the following: `apiMaj=10&apiMin=1&oneboxName=app&query=james`.

With a suggested content provider, set the URL template as `http://host.company.com/apps/directory.jsp?apiMaj=10&apiMin=1&oneboxName=app&query=$ora:q1`. The provider pattern is the same: `dir\s(\S+)`. The XSLT used for Google OneBox can be re-used with a minor change. Look for the line:

```
<xsl:template name="apps">
```

and change that line in your template to

```
<xsl:template match="/OneBoxResults">
```

Customizing the Appearance of Search Results

You can customize the default look and feel of the search result list for the default query application.

To customize the appearance of search results:

1. On the **Global Settings** tab, select **Configure Search Results List** under Out-of-Box Query Application.

    The Configure Search Results List page is displayed.

2. Select **Use Advanced Configuration**, then make the desired customizations.

3. Select attributes to appear in the XML description of result documents.

    The available attributes are local attributes, federated attributes, and internal attributes. Each attribute name appears only one time. That is, the name of a federated attribute with the same name as another attribute or with an explicit mapping to a local search attribute appears only once. Table 5–1 describes Oracle SES internal attributes.

4. Provide extensible style sheet language transformations (XSLT) to operate on the selected attributes. The default XSLT is installed at

    ses_home/search/webapp/defaults/advanced_config.xsl

    This XSLT transforms XML content into an HTML fragment to be displayed in the result list. If the XSLT is blank, then the untransformed result XML is displayed in the result list. You can use this output for debugging purposes while configuring the XSLT. Invalid XSLTs cannot be saved.

    The output of this transformation should be HTML. Enter the following in the XSLT:

    ```
    <xsl:output method="html" />
    ```

5. Provide a cascading style sheet (CSS) to style the HTML generated in the XSLT. The default CSS is installed at

    ses_home/search/webapp/defaults/advanced_config.css

    This CSS is used along with the included CSS files for the default query application. When there are conflicts between the advanced configuration CSS and the default CSS files, the advanced configuration definitions are used. Default XSLT and CSS style sheets are provided for Advanced Configuration.

Table 5–1   Oracle SES Internal Attributes

Name	Type	Description
eqcacheurl	String	The URL of the cached version of the document; that is, the value of the "Cached" link in the default result list.
eqcontentlength	Number	The size of the document in bytes.
eqdatasourcename	String	The (untranslated) name of the source where the document originated. This name is local to the instance that the document came from and not the instance that is receiving the document.
		If a document comes from a federated source, then the value of this attribute is the name of the source on the federated instance, and not the name of the federated source on the local instance.
eqdatasourcetype	String	The (untranslated) type of source where the document originated. This type is local to the instance from which the document came. For example, Federated is not a valid value for this attribute.
eqdocid	Number	Document ID.

Table 5–1 (Cont.) Oracle SES Internal Attributes

Name	Type	Description
eqfedchain	String	The chain of instance names representing the path of a federated document. The instance names are delimited by a semicolon (;).
eqfedid	String	The federated source ID chain delimited by an underscore (_).
eqgroupbrowseurl	String	The URL to browse the **infosource** group; that is, the value of the "Source Group" link in the default result list.
eqlinksurl	String	The URL of the page containing a list of links to the document; that is the value of the "Links" link in the default result list.
eqpathbrowseurl	String	The URL to browse the infosource path; that is, the value of the "Path" link in the default result list.
eqredirecturl	String	The redirect URL to the original document; that is, the value of the title link in the default result list.
eqsimilardoc	Boolean	A value of `true` indicates a similar document; otherwise, `false`.
eqsnippet	String	The excerpt or keyword in context of the document.
equserquery	String	The query string.

XML Result Schema

To apply the XSLT, each document is converted into an XML description at query-time with the following schema:

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
   <xsd:element name="result">
   <xsd:complexType>
   <xsd:all>
      <!-- Internal attributes -->
      <xsd:element name="eqdatasourcename" type="xsd:string" maxOccurs=1 />
      <xsd:element name="eqdatasourcetype" type="xsd:string" maxOccurs=1 />
      <xsd:element name="eqsnippet" type="xsd:string" maxOccurs=1 />
      <xsd:element name="eqredirecturl" type="xsd:string" maxOccurs=1 />
      <xsd:element name="eqcacheurl" type="xsd:string" maxOccurs=1 />
      <xsd:element name="eqlinksurl" type="xsd:string" maxOccurs=1 />
      <xsd:element name="eqsimilardoc" type="xsd:boolean" maxOccurs=1 />
      <xsd:element name="eqcontentlength" type="xsd:integer" maxOccurs=1 />
      <xsd:element name="equserquery" type="xsd:string" maxOccurs=1 />
      <xsd:element name="eqgroupbrowseurl" type="xsd:string" maxOccurs=1 />
      <xsd:element name="eqpathbrowseurl" type="xsd:string" maxOccurs=1 />
      <xsd:element name="eqdocid" type="xsd:integer" maxOccurs=1 />
      <xsd:element name="eqfedid" type="xsd:string" maxOccurs=1 />
      <!-- Built-in search attributes -->
      <xsd:element name="author" type="xsd:string" maxOccurs=1 />
      <xsd:element name="description" type="xsd:string" maxOccurs=1 />
      <xsd:element name="headline1" type="xsd:string" maxOccurs=1 />
      <xsd:element name="headline2" type="xsd:string" maxOccurs=1 />
      <xsd:element name="headline3" type="xsd:string" maxOccurs=1 />
      <xsd:element name="host" type="xsd:string" maxOccurs=1 />
      <xsd:element name="infosource" type="xsd:string" maxOccurs=1 />
      <xsd:element name="infosourcepath" type="xsd:string" maxOccurs=1 />
      <xsd:element name="keywords" type="xsd:string" maxOccurs=1 />
      <xsd:element name="language" type="xsd:string" maxOccurs=1 />
```

```
                <xsd:element name="lastmodifieddate" type="xsd:date" maxOccurs=1 />
                <xsd:element name="mimetype" type="xsd:string" maxOccurs=1 />
                <xsd:element name="referencetext" type="xsd:string" maxOccurs=1 />
                <xsd:element name="subject" type="xsd:string" maxOccurs=1 />
                <xsd:element name="title" type="xsd:string" maxOccurs=1 />
                <xsd:element name="url" type="xsd:string" maxOccurs=1 />
                <xsd:element name="urldepth" type="xsd:integer" maxOccurs=1 />
                <!-- Custom search attributes -->
                        .
                        .
                        .
        </xsd:all>
        </xsd:complexType>
        </xsd:element>
</xsd:schema>
```

XML has the following rules for element names:

- Alphanumeric and non-English characters, numbers, and ideograms, are allowed

- Limited punctuation is allowed: underscore, hyphen, and period

- Names can only begin with letters, ideograms, and underscores

Custom attribute names must conform to these rules for advanced result rendering. To enforce these rules, the empty string replaces all characters that are not permitted by these rules. In addition, Oracle SES search attributes are case-insensitive, and therefore all attributes are converted to lowercase when used in XML format.

For example, suppose the raw XML result data is as follows.

```
<result>
   <eqdatasourcetype>WEB</eqdatasourcetype>
   <title>Oracle Secure Enterprise Search</title>
   <url>
      http://www.oracle.com/technetwork/search/oses/overview/index.html
   </url>
      <author>Anonymous</author>
   <description>
      Oracle Secure Enterprise Search 10g, a standalone product from Oracle,
enables a secure, high quality, easy-to-use search across all enterprise
information assets.
   </description>
 </result>
```

The following XSLT extracts and formats the title, URL, and author for documents coming from Web sources:

```
<?xml version="1.0" encoding="UTF-8" ?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
 <xsl:template match="result[eqdatasourcetype='WEB']">
   <span class="title">
     <xsl:text> &quot;</xsl:text><xsl:value-of select="title" />
     <xsl:text>&quot;</xsl:text>
   </span>
   <span class="author">
     <xsl:text> By </xsl:text><xsl:value-of select="author" />
   </span>
   <br/>
   <span class="url">
     <a href="http://{url}"><xsl:value-of select="url" /></a>
   </span>
```

```
 </xsl:template>
</xsl:stylesheet>
```

A CSS style sheet for this output may be:

```
.title { font-weight: bold; }
.url { font-style: italic; }
```

These style sheets produce a final result of:

"Oracle Secure Enterprise Search" By Anonymous

http://www.oracle.com/technetwork/search/oses/overview/index.html

Configuring and Using Clustering in Search Results

Real-time clustering dynamically organizes search results into groups to provide end users with different views on the top results. Clustered documents within one group, called a cluster node, share the same common topics or property values. A cluster node with a large document set can be categorized into child cluster nodes, and a hierarchy is built. Users can navigate directly to a specific cluster node. Effective real-time clustering balances clustering quality and clustering time.

Search attributes (String, Number or Date) are used to generate a cluster tree. The attributes can be local search attributes, federated attributes that are not explicitly mapped, and Oracle SES internal attributes.

Oracle SES supports two types of cluster trees: topic and metadata. Each tree can be enabled or disabled individually. Parameters that apply to all cluster trees for the default query application can be configured on the Global Settings - Clustering Configuration page. These include the following:

- **Enable clustering**: Select this option to enable clustering.

- **Maximum cluster tree depth**: The maximum level of the cluster node hierarchy.

- **Maximum number of children per node**: The maximum number of cluster nodes on each level. This does not apply to the miscellaneous node.

- **Minimum number of documents per node**: The minimum number of the documents within one node. This does not apply to the miscellaneous node.

  Within each level of a cluster tree, documents that are not categorized into a node are placed in a special node named `miscellaneous`. The **Minimum number of documents per node** and **Maximum number of children per node** parameters do not apply to the miscellaneous node.

For customized Oracle SES applications, configure clustering with the Query Web Services API.

> **See Also:** "Oracle Secure Enterprise Search Web Services APIs" on page 13-2

Topic Clustering

Topic clustering uses the most significant phrases (and optionally sentences) from documents to create relevant cluster nodes and hierarchies. The significant phrases are extracted both at query-time and by the Secure Enterprise Search Document Summarizer, which is a document service included by default for search result clustering.

Configure crawl-time extraction of top phrases with document services parameters on the Global Settings - Document Services page. Create a topic clustering tree on the Global Settings - Clustering Configuration - Create Topic Clustering Tree page.

Topic clustering can be configured with one or more search attributes of String type and with the following Oracle SES internal attributes:

- `eqsnippet`: The excerpt of the document with keywords in context.

- `eqtopphrases`: The most frequent phrases within one document among the phrases with the same number of words.

- `eqtopsentences`: The significant sentences within one document based on the significant phrases.

By default, the attributes `keywords`, `title`, `eqsnippet` and `eqtopphrases` are configured for topic clustering. `Keywords`, `eqtopphrases`, and `eqtopsentences` contain pre-extracted words and phrases: no additional phrase extraction is performed on these attributes.

Parameters that control query-time word and phrase extraction for the default query application can be configured on the Global Settings - Clustering Configure page. These include the following:

Single Word Extraction

- **Minimum occurrence**: The minimum frequency for the word to be extracted.

- **Maximum number of words to extract**: The maximum number of words to be extracted.

Phrase Extraction

- **Minimum occurrence**: Minimum frequency for a phrase to be extracted.

- **Maximum number of phrases to extract**: Maximum number of phrases to be extracted.

- **Maximum phrase length**: Maximum number of words for each phrase to be extracted.

Topic clustering uses a phrase stop words list and a blacklist to prevent words or phrases from becoming topic cluster result nodes.

The phrase stop words list is also used by the Document Summarizer document service. The stop words file is a language-specific file containing words that should not be considered during phrase extraction. The blacklist file is a language-specific file containing words and phrases that should not appear as cluster node names.

For example, if all indexed documents include the phrase "Oracle Corporation" and it does not make sense to have a cluster node for "oracle corporation", then this phrase could be added to the blacklist.

> **Note:** A separate stop words list contains index stop words. This is an Oracle SES internal file for words that should not be indexed. This list is not related to phrase extraction.

Both the stop words and the blacklist files are in plain text format, with each line of text containing one word or phrase. The phrase stop words file name should be "phrasestopwords" followed by period and two-letter language code (for example, `phrasestopwords.en` for English). Similarly, the blacklist file name should be "blacklist" followed by period and two-letter language code.

By default, these files are located in the directory:

ses_home/search/lib/plugins/doc/extractor/phrasestopwords

Sample phrase stop words files for other languages (non-English) are in the directory:

ses_home/search/lib/plugins/doc/extractor/samples/phrasestopwords

If there are documents in any of the non-English languages, then copy the related language files from the above mentioned directory to the default stop words directory:

ses_home/search/lib/plugins/doc/extractor/phrasestopwords

The order of words and phrases in the stop words file does not affect the phrase extraction. The following is a sample `phrasestopwords.en` file:

```
a
an
me
:
z
```

The following is a sample `blacklist.en` file:

```
site maps
oracle corporation
:
term of use
```

> **Note:** The stop words and the blacklist are used by both the default query application and the Web services API.

Topic clustering currently works best in English. Both the document summarizer in the crawler and the clustering module in the query application use a **stemmer** to stem the word and merge the words and phrases with the same stems. The open source stemmer library Snowball is used for this purpose. The version included with Oracle SES supports the following languages:

- Dutch
- English
- Finnish
- French
- German
- Norwegian
- Portuguese
- Russian
- Spanish
- Swedish

The Egothor stemmer is included for Polish language support. The stemmer configuration is shared between the default query application and the Web Services API.

> **Note:** Topic clustering is not supported for Chinese and Japanese.

> **See Also:** Appendix C, "Third Party Licenses"

Metadata Clustering

Metadata clustering is performed on a single attribute of String, Date, or Number type. If there are multiple values for the same attribute in one document, then only the first value is used for clustering. By default, the entire value is passed in as is for clustering.

However, for String attributes only, a delimiter can be specified for tokenizing the attribute value. If no tokenization delimiter is entered (or if only white space is entered), then the delimiter defaults to white space. When tokenized, the single attribute value is divided into multiple segments and each segment can correspond to a hierarchy based on another delimiter called the hierarchy delimiter. White space is the default hierarchy delimiter; however, if both tokenization and hierarchy are selected, then the delimiters must be different. Parsing is done first by tokenization, and then by interpreting the hierarchy from the resulting tokens.

Create a metadata clustering tree on the Global Settings - Clustering Configuration - Create Metadata Clustering Tree page.

As an example where both tokenization and hierarchy are meaningful, a category attribute might consist of a comma-delimited list of fields, each representing a slash-separated hierarchical categorization (as in "java/j2ee/jdbc, oracle/search/connector").

The tokenization and hierarchy configuration is not applicable to Date or Number attributes. Metadata trees of Date type attributes use a fixed display format with year on the first level, month on the second, and day on the third. The year is sorted in descending order, and the month and day are sorted in ascending order.

Metadata trees for Number type attributes are range-based with a fixed number of ranges (5) and a fixed tree depth (3), that is, the maximum number of ranges for number clustering trees is five (5). The tree depth starts at the root node. For a range to be shown, it must satisfy the Minimum Documents Per Node parameter, which is set on the Query-time Clustering Configuration page. Empty ranges are not shown.

> **See Also:**
>
> - "Configuring Clustering in the Web Services API" on page 5-11
>
> - "Document Service API" on page 13-44
>
> - "Customizing the Appearance of Search Results" on page 5-3 for descriptions of Oracle SES internal attributes
>
> - "Oracle SES Stoplist" on page 4-29

Using Clustering

Cluster nodes filter the top results but do not change the order of the documents. When users select a cluster node, the result view is limited to the documents in that cluster node. All operations, such as sorting or paging through results, are limited to the cluster node.

The real-time clustering sidebar is hidden by default. Users can display the sidebar by clicking an arrow icon on the left-hand side of the search results page. Within the sidebar, result clusters are shown. The cluster nodes are sorted by the number of documents in each node.

Users can expand or collapse the nodes within a cluster tree without affecting the rest of the interface. If users click a cluster node, then the search results are filtered. If a cluster tree contains no children nodes, it is disabled.

Configuring Clustering in the Web Services API

Methods in the Query Web Service API provide clustering for customized Oracle SES applications. The main interface is the method `doOracleOrganizedSearch`, which accepts query information, grouping and sorting options, and clustering requests. Based on the request variation, it returns the requested result. A second method `doOracleFetchSearch` is used when the set of documents is known.

The input for `doOracleOrganizedSearch` includes the following information

- Query

- TopN (the result set size used for grouping, sorting, and clustering)

- Duplicate controls (removed, marked)

- Data group list

- Query and document language

- Grouping and sorting options

- Cluster tree configuration info (tree depth, children for each node, threshold, tree format type: JSON, XML; topic extraction configuration, metadata clustering configuration.)

- Other query parameters (including Number startIndex, Boolean returnCount, String filterConnect, Filter[] filters)

The output is an object that contains the search result, grouping information, and the cluster tree string list. The search result list is in the order specified by the grouping and sorting option. If this is not specified, then it is sorted by the relevance score. The returned cluster tree string represents the clustering tree information: tree structure, node names, and document IDs.

> **See Also:** "Search Operations" on page 13-12

Java Classes for Clustering

There are three classes to support the grouping and sorting options: `GroupAttribute`, `SortAttribute`, and `GroupResult`.

There are two classes to support the clustering request: `ClusterConfig`, which controls the clustering request, and `ClusterTree`, which contains the tree output.

The class `OracleResultContainer` is defined to wrap the search hit result, grouping result, and clustering result.

`doOracleFetchSearch` is used for fetching a selected list of documents identified by their document ID, federated source ID, or both.

If `GroupAttribute` is specified, then it is automatically added to the top of the sorting attribute. For example, if the query is grouped by host name and sorted by title, then the search hit is sorted by (hostname, title).

The sorting, grouping, or clustering option can be applied to this result. Sorting is based on the top N result, while grouping and clustering is based on the result window determined by (`startIndex`, `docsRequested`).

Cluster Result XML Schema

The main XML element, node, contains the following attributes:

- `id`: ID for the node. The value represents the full path with the parent node paths.

- `name`: The name of the node. This is actually the topic for the node.

- `level`: The cluster node level started from 1 for the top node.

- `size`: Number of documents under (directly and indirectly) this cluster node.

- `leaf`: This is "1" if the cluster node only contains documents and no child cluster nodes. Otherwise, this is "0".

- `keywords`: All keywords and phrases within the cluster node.

The node element contains the document IDs in the XML text element if the node is a simple node. The document ID in the XML file has the format `docID.SES_InstanceID`. If the document is from the local instance, then the `SES_instance_ID` is omitted.

```
<cluster>
   <nodeset>
      <node id="1" name="all" level="1" size="100" leaf="0" keywords="all"/>
      <node id="1.4" name="java" level="2" size="99" leaf="0" keywords="java"/>
      <node id="1.4.1" name="data warehousing" level="3" size="38" leaf="0"
         keywords="technologies bi,data warehousing,linux .net office
            php security service"/>
      <node id="1.4.1.1" name="tutorials blogs" level="4" size="12" leaf="1"
         keywords="tutorials blogs">
         2773,8031,109,8033,806,26940,817,8024,8030,2862,8032,8028
      </node>
      <node id="1.4.1.2" name="stored procedure" level="4" size="4" leaf="1"
         keywords="stored procedure">
         4239,4243,2784,4335
      </node>
      <node id="1.4.1.3" name="miscellaneous" level="4"  size="22" leaf="1">
         4017,2836,8029,2767,1502,113814,11731,1138,392,2819,2763,1421,
         221,705,7739,2838,2749,2351,2802,1158,15751,15747
      </node>
   </nodeset>
</cluster>
```

Cluster Result JSON Format

To integrate with AJAX applications, the cluster results can be returned in JSON format. The JSON format directly reflects the tree structure of the cluster results. Each node has a child array, which is a list of nodes representing the direct children of that node, or a docs array representing the document in that node if the node is a leaf node. Nodes in the child array may have children, and so on.

Here is sample JSON output.

```
{"nodeset":

  {"id":"1",
  "name":"all",
  "level":1,
  "size":100,
  "leaf":false,
  "keywords":"all",
  "children":
     [{"id":"1.4",
```

```
        "name":"java",
        "level":2,
        "size":99,
        "leaf":false,
        "keywords":"java",
        "children":
            [{"id":"1.4.1",
            "name":"data warehousing",
            "level":3,
            "size":38,
            "leaf":false,
            "keywords":"technologies bi,data warehousing,linux .net office php
security service",
            "children":
                [{"id":"1.4.1.1",
                "name":"tutorials blogs",
                "level":4,
                "size":12,
                "leaf":true,
                "keywords":"tutorials blogs",
"docs":["2773","8031","10","803","806","26940","817","8024","8030","2862","803","8
028"] },
                {"id":"1.4.1.2",
                "name":"stored procedure",
                "level":4,
                "size":4,
                "leaf":true,
                "keywords":"stored procedure",
                "docs":["4239","4243","2784","4335"]}]
            }]
        },
        {"id":"1.5",
        "name":"miscellaneous",
        "level":2,
        "size":1,
        "leaf":true,
        "docs":["265915"]
        }]
    }
}
```

Configuring and Using Facets in Search Results

The faceted search provides the mechanism for filtering data in the search result based on the categorization and the sub-categorization of the data. The various categories of data can be considered as facets, and for each category, the sub-categories are called facet nodes. A sub-category again can have sub-sub-categories, that is, a facet node can have child nodes, and so on. A complete node hierarchy of a facet is called *facet tree*.

For example, all the books in a book store can be categorized based on the subject of the books, such as, history, art, literature, science, and so on. The science books can be again categorized into physics, chemistry, biology, and so on.

In the faceted search, books as a whole can be considered as a facet. Books can be broadly categorized based on a subject, such as, art, literature, science, and so on, and are considered as the root nodes of the *books* facet. The science category of books can be further sub-categorized into physics, chemistry, biology, and so on, and are considered as the child nodes of the *science* node. Thus, the user can refine the search result for the

books facet, based on the *science* node, and further based on its child nodes, such as *physics*, *chemistry*, *biology*, and so on.

In Oracle SES, facets can be created based on the search attributes of String, Number, and Date data types.

Using Facets in Query Application

In the Oracle SES query application, the search result page shows the following facet navigation components that can be used for filtering the search result:

Facet Panel

A facet panel is a container for facets that displays all the facets configured in Oracle SES. A facet is a tree structure containing multiple facet nodes. For example, *Books*, *Computer*, and *Location* are various facets. The facet named *Location* can have various facet nodes, such as, *Los Angeles*, *New York*, and *New Jersey*. The facet named *Computer* can have various facet nodes, such as, *Desktop*, *Laptop*, and *Netbook*. You can select facet nodes from the facet panel to use it as a search criteria for filtering a search result.

Facet Navigation Bar

A facet navigation bar displays the complete facet node hierarchy currently selected by the user from the facet panel. It also shows the query term currently entered in the search text box. The search result is filtered based on the selected facet node hierarchy. Each selected facet node is used as a search criteria for filtering the search result.

Using Facets in Query API

Oracle SES query API provides the following methods for faceted search:

- `doOracleFacetSearch`: This method returns faceted search results for a query. See "doOracleFacetSearch Message" on page 13-18 for more information.

- `getFacetNodes`: This method returns the facet child nodes for the specified facets having non-zero document count. See "getFacetNodes Message" on page 13-24 for more information.

Configuring Facets using Administration GUI

The Oracle SES Administration GUI provides the following pages for configuring facets:

- **Global Settings - Facets**

  This page lists all the facets defined in Oracle SES. Use this page to create, edit, and delete facets and facet nodes.

  > **Note:** Oracle SES supports facets of String, Number, and Date types.

- **Global Settings - Translate Facet Name**

  Use this page to translate facet names and facet node names in different languages.

  The language to which the facet name and facet node names need to be translated must be specified using two-letter language code adhering to ISO 639-1 standard, except for Chinese and Portuguese languages (use `zh_CN` for simplified Chinese, `zh_TW` for traditional Chinese, and `pt_BR` for Portuguese, Brazilian). Default is `en`, that is, English.

■ **Global Settings - Configure Facets**

Use this page to enable or disable facets, and configure various ways to display facets in the query application, such as, ordering of facets, number of facets to display, number of facet nodes to display for each facet, ordering of facet nodes for each facet, and so on.

■ **Global Settings - Configure Source Groups**

Use this page to configure about which facets to show in the query application page for a particular source group.

Configuring Facets using Administration API

The following Administration APIs can be used for configuring facets in Oracle SES:

■ **facetTree**: Use this object to create, update, delete, export, and translate facets and facet nodes.

■ **queryUIFacets**: Use this object to enable or disable facets, and to configure various display properties of facets, such as number of facets to display, sort order of facets, number of facet nodes to display for each facet, and sort order of facet nodes for each facet.

■ **queryUISourceGroups**: Use this object to configure about which facets to show in the query application page for a particular source group.

> **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for more information about the Administration APIs for configuring facets.

Faceted Search in Federation Environment

A federation provides a unified framework to search different repositories that are crawled, indexed, and maintained separately. In a federation environment, Oracle SES aggregates the facet trees in a search result from all the Oracle SES instances, to show a unified facet search result. This is achieved using the concept of federation *broker* and federation *endpoint*. The facet nodes returned from all the Oracle SES instances (federation endpoints) in a federation are merged at the federation broker. Thus, in a faceted search in a federation environment:

■ the document count of the merged facet node is the sum of the document counts of all the facet nodes from all the federation endpoints.

■ the facet node names of all the merged facet nodes is the union of all the facet node names from all the federation endpoints.

■ the document count for the merged facet node is the sum of the document counts of all the facet nodes from all the federation endpoints that have the same facet node name.

> **Note:** For the faceted search to work in a federation environment:
>
> - facets defined at all the endpoints in a federation must also be defined at the broker.
>
> - a facet defined at a broker and at all the endpoints in a federation, must have the same facet tree paths, the same facet names and facet node names, and the same facet node values, that is, a facet defined in a federation environment must have the same configuration across all the Oracle SES instances. The facet node names are treated as case sensitive, for example, the facet node `SMALLNUM(60-1000)` defined at a broker and the facet node `smallNum(60-1000)` defined at an endpoint are treated as different facet nodes.
>
> - the infosource search functionality must be disabled, that is, infosource search and faceted search cannot be used together in a federation environment.

Faceted Search Result Count versus Normal Search Result Count

Oracle SES default query application shows search result count on the upper-right corner of the search result page. When faceted navigation is enabled, the facet tree nodes also show document count for each of the nodes. In some scenarios, these two counts will be different for the same query. The following are some of the main scenarios when the document count of the normal query search result will be different from the document count of the faceted query search result.

> **Note:** The search result count in the Oracle SES query application by default is calculated based on the **Approximate count** setting of the **Hit Count Method** parameter present on the **Global Settings - Query Configuration** page of the Administration GUI. When the faceted search is used, the search result count is calculated based on the **Exact count** setting of the **Hit Count Method** parameter.

Scenario 1

The faceted search result count supports multiple values for a given facet. It is possible that a document can be counted more than once for a faceted search result count, and in that case, the faceted search result count will be different from the normal search result count. For example, if *author* is a facet and there are two authors for a given document, then for that document the author facet search result count will be two, with one count for each author, whereas the normal search result count for that document will be one.

Scenario 2

The Oracle SES query application handles similar and duplicate documents based on the value of the **Similar Document Handling** parameter present on the **Global Settings - Query UI Configuration** page of the Administration GUI. When this parameter value is set to **Detect**, then on the normal search result page, duplicate documents are not shown, and similar documents are grouped together and are displayed only when the **Similar Documents** link is clicked. When this parameter value is set to **Remove**, then on the normal search result page, duplicate documents are not shown, and a single document is shown in place of multiple similar documents.

Thus, when the value of the **Similar Document Handling** parameter is either **Detect** or **Remove**, the normal search result count will exclude similar and duplicate documents. But, the faceted search result count will always include similar and duplicate documents. This will lead to the discrepancy between the normal search result count and the faceted search result count. To avoid this discrepancy, the value of the **Similar Document Handling** parameter should be set to **Disabled**, but in that case, similar as well as duplicate documents will be displayed in the normal search result.

Configuring and Using Tags in Search Results

By assigning tags to a document, you can classify a document into multiple categories. This enables searching relevant documents based on certain tags or categories. For example, a company's quarterly earnings report can be tagged as "financial report" or "company report". So, when you search using any of these two tags as query terms, you will get the company's quarterly earnings report as one of the top-N results.

Configuring Tags using Administration GUI

You can specify the following configuration parameters in the **Search - Tagging** page of Administration GUI related to the tagging functionality:

Tagging Mode

Select **Enabled** to enable the tagging functionality. You can enable the tagging functionality for all the users (anonymous users), or for a group of users, or only for all the logged-in users.

> **Note:**
>
> - All types of users (Oracle SES admin user, Oracle SES non-admin user, and anonymous user) can add tags to documents.
>
> - An admin user can remove all the tags.
>
> - A non-admin user can remove only those tags that are added by that user.
>
> - An anonymous user (user who is not logged-in) cannot delete tags.

Maximum Tags per Document

Maximum number of tags that can be assigned to a document (not specific to a user).

Maximum Tags per Session

Maximum number of tags that can be added in a session.

Cut-off (days)

Number of days for which any tag should be available in the query application, even if it is not being used. When the number of days specified for **Cut-off (days)** elapse, the tags that are unused for these number of days are removed from the query application.

Configuring Tags using Administration API

You can use the `tagging` and `tag` objects of Administration API to configure and upload tags in Oracle SES.

> **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for more information about the `tagging` and `tag` objects.

Configuring Tags using searchadminctl Tool

You can use the `searchadmin` command-line tool to perform the following operations related to tags:

Table 5–2   Tag operations using searchadmin tool

Operation	Syntax	Description

Bulk Upload Tags	`searchadmin createAll tag --INPUT_FILE=<XML file containing tags to upload>`	Uploads tags in bulk. Tags to upload are specified in an XML file.
		
		 Returns CREATE_SUCCEEDED for each tag that is successfully created.
		
		 Returns DUPLICATE_IGNORED for each tag that already exists.
		
		 Sample XML file for bulk upload:
		
		 `<?xml version="1.0"?>`
		`<search:config`
		`productVersion="11.2.2.2.0"`
		`xmlns:search="http://xmlns.oracle.com/search">`
		`  <search:tags>`
		`    <search:tag>`
		
		`      <search:name>`
		`        oses`
		`      </search:name>`
		
		`      <search:docUrl>`
		`        http://www.oracle.com/xyz.html`
		`      </search:docUrl>`
		
		`      <search:owner>`
		`        abc@oracle.com`
		`      </search:owner>`
		
		`    </search:tag>`
		`  </search:tags>`
		`</search:config>`
		
		 The following is the description for various tag elements:
		
		 ■ tag: Contains details for a tag.
		 ■ name: Name of a tag.
		 ■ docURL: Document to tag (in URL form).
		 ■ owner: Owner of the document.

Delete All Tags	`searchadmin deleteAll tag [-s]`	Deletes all tags. Use the -s option to ignore the invalid state errors during the delete operation. Returns DELETE_SUCCEEDED for each tag that is successfully deleted.

Table 5–2   (Cont.) Tag operations using searchadmin tool

Operation	Syntax	Description

Delete Specific Tags	`searchadmin deleteList tag -k <Text file containing tags to delete> [-s] [-f]`	Deletes specific tags.
		
		Use -f option to ignore error, if a tag to be deleted is not found.
		
		Returns NOT_FOUND_IGNORED, if a tag to be deleted is not found.
		
		Sample text file containing tags to delete:
		
		`--NAME=oses23 --DOC_`
		`URL=http://stanc17:7777/testdata/seedurls`
		`/duplicate/near.html --OWNER=public`
		
		`--NAME=oses24 --DOC_`
		`URL=http://stanc17:7777/testdata/seedurls`
		`/duplicate/identical.html --OWNER=public`

Export All Tags	`searchadmin exportAll tag -o <Text file to store exported tags>`	Exports all the tags.

Export Specific Tags	`searchadmin exportList tag -k <Text file containing tags to export> -o <Text file to store exported tags>`	Exports the tags specified in an input file (using -k parameter). The input file can contain special operators to export tags based on regular expressions.
		
		Sample text file containing tags to export (this file is similar to the delete tags file):
		
		`--NAME=oses23 --DOC_`
		`URL=http://stanc17:7777/testdata/seedurls`
		`/duplicate/near.html --OWNER=public`
		
		`--NAME=oses24 --DOC_`
		`URL=http://stanc17:7777/testdata/seedurls`
		`/duplicate/identical.html --OWNER=public`

Activate Tagging	`searchadmin activate tagging`	Enables the tagging functionality. Default tagging mode is set to allow all the logged-in users to tag results. If this command is executed when tagging is already activated, INVALID_STATE_IGNORED is returned.
Deactivate Tagging	`searchadmin deactivate tagging`	Disables the tagging functionality.

Table 5–2   (Cont.) Tag operations using searchadmin tool

Operation	Syntax	Description

Update Tagging Configuration	`searchadmin update tagging -i <XML file containing the tagging configuration>`	Updates the tagging configuration specified in an XML file.
		
		Returns UPDATE_SUCCEEDED, if the update operation is successful.
		
		If you try to deactivate tagging using update, INVALID_STATE_IGNORED is returned.
		
		Sample XML file containing tagging configuration:
		
		`<?xml version="1.0" encoding="UTF-8"?>`
		`<search:config`
		`productVersion="11.2.2.2.0"`
		`xmlns:search="http://xmlns.oracle.com/search">`
		`    <search:tagging>`
		`<search:maxTagPerDoc>100</search:maxTagPerDoc>`
		`<search:maxTagPerSession>1000</search:maxTagPerSession>`
		`<search:tagCleanupInterval>32</search:tagCleanupInterval>`
		`<search:authorizationMode><search:allUsers></search:authorizationMode>`
		`    </search:tagging>`
		`</search:config>`
		
		The following is the description for each tagging configuration:
		
		■ maxTagPerDoc: Maximum number of tags that can be assigned to a document (not specific to a user).
		
		■ maxTagPerSession: Maximum number of tags that can be added in a session.
		
		■ tagCleanupInterval: Number of days for which any tag should be available in the query application, even if it is not being used. When the number of days specified in tagCleanupInterval elapse, the tags that are unused for specified number of days are removed from Oracle SES.
		
		■ authorizationMode: This paraemeter controls who is allowed to use the tagging feature. The following are the valid values for the authorizationMode parameter:
		
		`disabled` - Tagging is disabled for all the users.
		
		`loggedInUsers` - Tagging is enabled only for the users who are logged-in.
		
		`authorizedPrincipals` - Tagging is enabled only for the specific users having tagging privilege specified using the `authorizedPrincipal` object.
		
		`allUsers` - Tagging is enabled for all the users (anonymous tagging).

Assigning Tags to Documents in Search Results Page

If the tagging feature is enabled, then a tag icon is displayed beside all the documents in the search result page, along with the tags that are already assigned to the documents. You can click the tag icon to assign or remove tags for a document.

Configuring Default Sort Conditions

The default sort conditions for search results can be configured in the **Default Sort Conditions** section on the **Global Settings - Query Configuration** page in the Administration GUI. In this page, multiple sort conditions are specified by specifying sortable attributes with their corresponding sort order, such as ascending or descending. The search results get sorted according to the order in which the sortable attributes are configured in this page.

The default sort conditions for search results for each source group can also be configured in the Administration GUI, in the **Absolute Sorting** tab on the **Configure Source Groups** page for a specific source group. In this page, you can either specify to use the same default sort conditions that are configured in the **Default Sort Conditions** section on the **Global Settings - Query Configuration** page, or specify different default sort conditions.

> **Note:** Source group specific default sort conditions specified on the **Configure Source Groups** page take precedence over the generic default sort conditions specified on the **Query Configuration** page.

Configuring Top-N Documents and Group/Sort Attributes

The default number of documents to retrieve for top-N results, the maximum number of documents to retrieve for top-N results, the groupable attributes, and the sortable attributes can be configured either in the Administration GUI, on the **Global Settings - Query UI Configuration** page, or using the Administration API object **queryUIConfig**.

Top-N Documents

The default top-N documents setting **Default Number of Results** on the **Query UI Configuration** page represents the number of documents retrieved by default as part of the AJAX call for result clustering, grouping, and sorting.

To page through a very large result set, say 500 documents, the user may view a page of results beyond the default top-N value. Suppose top-N is set to the default 100, and the user wants to view the results numbered 150-160. To provide result clustering and sorting/grouping, the browser must request 160 results. If the user views page 490-500, then the browser would be requesting 500 results through the AJAX call. This may result in reduced performance.

The maximum top-N documents setting **Maximum Number of Results** on the **Query UI Configuration** page represents a threshold above which the query application displays only a single page of results.

This mode does not provide any sorting, grouping, or result clustering. However, it lets a user to view the entire result set without the costly subsequent retrievals of top-N results.

Suppose maximum top-N is 200. If a user is viewing results 30-40, then the browser would retrieve the default of 100 results. If the user views results 170-180, then the

browser would request 180 documents. If the user views results above 200, then the query application would display only the current page of results.

> **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for more information about the Administration API object `queryUIConfig` for specifying default top-N documents and maximum top-N documents for search results.

Group By and Sort By Lists

Oracle SES query application displays a set of attributes by default in the **Group by** and **Sort by** lists.

Default grouping attributes:

- (none)
- Author
- File Format
- Source
- Date

Default sorting attributes:

- Author
- File Format
- Title
- Relevance
- Path
- Language
- Date

You can also use the **Global Settings - Configure Source Groups** page of Administration GUI to configure source group specific set of sortable attributes to display in the **Group by** and **Sort by** lists. Refer to "Configuring and Using Source Groups in Search Results" on page 23 for more information.

> **See Also:** "Searching on Date Attributes" on page 3-9.

Configuring Sort Criteria using Sortable Attributes (Absolute Sort)

You can define any of the Oracle SES attributes as sortable attributes. You can use these sortable attributes to define a global sort criteria as well as source group specific sort criteria. The sortable attributes can also be used in the Sort by list of the query application for performing global sort as well as source group specific sort.

Configuring Sortable Attributes for Search Criteria in Administration GUI

You can define the search attributes in the **Global Settings - Search Attributes** page. In this page you can specify whether the search attribute should be used as a sortable attribute.

You can use the sortable attributes for defining the default sort criteria for search results in the **Default Sort Conditions** section of the **Global Settings - Query**

Configuration page. Apart from the sortable attributes, you can also specify **Relevance (System)** and **Absolute Date (System)** as part of the sort criteria.

You can specify the sortable attributes to be available in the Sort by list of the query application for Absolute sort, by configuring them in the **Absolute Sorting** section of the **Global Settings - Query UI Configuration** page. If at least one sortable attribute is specified, then by default, **Relevance** is also added to the Sort by list for Absolute sort.

You can also use the sortable attributes for defining source group specific sort criteria:

- Default sort criteria: This can be done by updating the **Default Sort Conditions** section in the **Absolute Sorting** tab of the **Edit Source Group** page.

- Sort by list sort criteria: This can be done by updating the **Sortable Attributes** section in the **Absolute Sorting** tab of the **Edit Source Group** page.

Configuring Sortable Attributes for Search Criteria in Administration API

The following are the Administration APIs where you can specify the sortable attributes for using them as a sort criteria for a query:

- **relevanceRanking** object: You can specify a default sort criteria for the search results using the **sortCondition** element of the **relevanceRanking** object.

- **queryUIConfig** object: You can specify the sortable attributes to show in the Sort by list box of the query application by using the **absoluteSorting** element of the **queryUIConfig** object.

- **queryUISourceGroups** object: You can specify the source group specific default sort criteria as well as source group specific Sort by list sort criteria for the search results in the query application by using the **defaultSortCondition** element and the **sortabeAttrs** element of the **queryUISourceGroups** object.

    > **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for more information about the Administration APIs for using sortable attributes as search criteria.

Configuring and Using Source Groups in Search Results

You can use the **Global Settings - Query UI Configuration** page of Administration GUI to specify the source groups to display in the query application.

You can also use the **Global Settings - Configure Source Groups** page of Administration GUI to configure the display properties of facets, cluster trees, top-n sortable attributes, top-n groupable attributes, and Absolute sorting criteria (default sort conditions and sortable attributes) for a source group that is enabled to be displayed in the query application.

The following are the various configuration settings on the **Global Settings - Configure Source Groups** page.

Facets

You can specify particular facets or all the facets to display in the query application for a source group.

Cluster Trees

You can specify particular cluster trees or all the cluster trees to display in the query application for a source group.

Top-N Sortable Attributes

You can specify particular sortable attributes or all the default sortable attributes to display in the **Sort by** list in the query application for a source group.

Top-N Groupable Attributes

You can specify particular groupable attributes or all the default groupable attributes to display in the **Group by** list in the query application for a source group.

Absolute Sorting

You can specify the default sort conditions to apply on search results for a source group. When multiple sort conditions are specified, the search results are ordered according to the order in which the sortable attributes are specified in the configuration page.

You can either specify the same default sort conditions that are configured in the Query Configuration page, or specify new sort conditions by adding multiple sortable attributes from the list of sortable attributes. The sortable attributes list also contains **Relevance (System)** to allow sorting based on relevance, and **Absolute Date (System)** to allow sorting based on the last modified date of the documents. You can also specify the order of sorting for each sortable attribute, that is, either ascending or descending.

You can also specify the sortable attributes to be displayed in the **Sort by** list on the query page for a source group for Absolute sorting. You can either specify the same sortable attributes that are configured in the Query UI Configuration page, or specify new sort conditions by selecting multiple sortable attributes from the list of available sortable attributes. When new sort conditions are defined, the first attribute from the sort conditions for absolute sort and **Relevance** is displayed in the **Sort by** list on the query page. If no sortable attributes are specified, the **Sort by** list for absolute sort is not displayed on the query page.

> **Note:** The **Sort by** list that is displayed for Absolute sorting is different from the **Sort by** list that is displayed for the Top-N sorting on the query page. Both these lists have the same name, that is, **Sort by**.
>
> The **Sort by** list that is displayed near to the query box is for Absolute sorting, that is, it allows you to sort all the entries in a search result.
>
> The **Sort by** list that is displayed just above the search results (and beside **Group by** list) is for Top-N sorting, that is, it allows you to sort only the top N entries in a search result.

Customizing the Relevancy of Search Results

You can customize the default Oracle SES ranking to create a more relevant search result list for your enterprise. Ranking is determined by default and custom attributes. Default attributes include title, keywords, description, and others. Different weights indicate the importance of each attribute for document relevancy. For example, Oracle SES gives more weight to titles than to keywords.

To customize the relevancy of search results, you can use the Administration API or the Query Web Services API.

See Also:

- The `relevanceRanking` object in the *Oracle Secure Enterprise Search Administration API Guide*

- "Query Web Services Example: Customizing Relevancy" on page 13-34

Using Relevancy Boosting for URLs and Query Terms

Relevancy boosting lets the administrator influence the order of documents in the result list for a particular search. You might want to override the default results for the following reasons:

- For a highly popular search, direct users to the best results

- For a search that returns no results, direct users to some results

- For a search that has no click-throughs, direct users to better results

In a search, each result is assigned a score that indicates how relevant the result is to the search; that is, how good a result it is. Sometimes you know the documents that are highly relevant to some search. For example, your company Web site could have a home page for XML (`http://example.com/XML-is-great.htm`), which you want to appear high in the results of any search for XML. You would boost the score of the XML home page to 100 for an XML search.

To boost the relevancy of search results for URLs and query terms, you can do the required configuration using::

- the Relevancy page in the Search tab of Administration GUI

  > **Note:**   The Locate by Search operation on the Relevancy page works only with the unsecured sources.

- the `boostedUrl` object of Administration API

  > **See Also:**   The `boostedUrl` object in the *Oracle Secure Enterprise Search Administration API Guide*

Part II

Creating Data Sources

This part describes the data sources that you can set up for Oracle SES to crawl and index. It contains the following chapters:

- Chapter 6, "Configuring Access to Built-in Sources"
- Chapter 7, "Configuring Access to Content Management Sources"
- Chapter 8, "Configuring Access to Collaboration Sources"
- Chapter 9, "Configuring Access to Applications Sources"

6

Configuring Access to Built-in Sources

Among the built-in sources are the data repositories familiar to everyone, such as files, Web sites, and e-mail. Most of them can be set up very quickly. This chapter contains the following topics:

- Setting Up Web Sources
- Setting Up Table Sources
- Setting Up File Sources
- Setting Up E-Mail Sources
- Setting Up Mailing List Sources
- Setting Up OracleAS Portal Sources
- Setting Up Federated Sources

Setting Up Web Sources

A Web source enables users to search a Web site. The following procedures identify the basic steps for setting up a Web source using the Oracle SES Administration GUI. For more information on each page, click **Help**.

Oracle SES is configured to crawl Web sites on the intranet within the corporate fire wall. To crawl Web sites on the Internet (external Web sites), Oracle SES requires the HTTP proxy server information. See the Global Settings - Proxy Settings page.

You should review the default crawling parameters before you start crawling Internet sources.

To create a Web source:

1. On the home page, select the **Sources** secondary tab to display the Sources page.

2. For Source Type, select **Web**.

3. Click **Create** to display the Create Web Source page.

4. Complete the following fields:

    - **Source Name**: Name that you assign to this Web source.

    - **Starting URLs**: The HTTP or HTTPS address of the Web site, starting at the top page to be searched.

    - **Self Service**: **Disabled** to use an identity management system or **Enabled** to prompt users for their credentials.

- **Start Crawling Immediately**: Select this option to accept the default parameters and begin crawling, or deselect it to defer crawling.

5. Click **Create** or **Create & Customize**.

6. Follow the steps for crawling and indexing a source in "Getting Started Basics for the Oracle SES Administration GUI" on page 2-1.

Figure 6–1 shows the Create Web Source page.

Figure 6–1   Creating a Web Source

To customize a Web source:

1. When creating a Web source, click **Create & Customize** on the Create Web Source page to display the Customize Web Source page.

   or

   After creating a source, click the **Edit** icon on the Home - Sources page.

2. Click the following subtabs and make the desired changes.

   - **Basic Settings**: The choices entered on the Create Web Source page.

   - **Boundary Rules**: Contents of a URL that include or exclude a page from crawling.

   - **Document Types**: Common document and image types that you can include or exclude from crawling. By default, Oracle SES crawls HTML, Excel, PowerPoint, Word, PDF and plain text.

   - **Authentication**: Configuration of HTTP, HTML forms, or Oracle Single-Sign-On methods of authentication. By default, no authentication is required.

   - **Authorization**: Configuration of an Access Control List or an authorization manager plug-in.

   - **Metatag Mappings**: Maps document attributes to Oracle SES search attributes. See "Web Document Attributes" on page 6-4.

   - **Crawling Parameters**: Sets a variety of crawling conditions, such as depth, language, HTTP cookies.

3. Click **Apply**.

Figure 6–2 shows the Customize Web Source page.

Figure 6–2   Customizing a Web Source

Boundary Rules for Web Sources

When creating a Web source, the host name of the seed (top level URL) is automatically added to the boundary rule. However, subsequent changes to the seed URL are not reflected automatically to the rule. Remember to synchronize the boundary rule if there is any change to the seed URL. Currently, Oracle SES does not remove crawled URLs even if the original seed is removed: everything is controlled by the boundary rules.

Authentication Configuration for Protected Web Sources

Oracle SES supports crawling of protected Web sources. This functionality can be achieved by configuring the authentication parameters in the **Authentication** tab in the Create or Edit page for the protected Web source in the Administration GUI. You can configure the authentication parameters for any one of the following authentication mechanisms for accessing the protected Web sources:

- **HTTP:** For HTTP authentication of basic and digest type, specify the proxy server details, such as, host name, realm, user name, and password.

- **HTML Forms:** For HTML form authentication, specify the address of the registered HTML form, form action, and success URL.

- **Single Sign-On:** For Single Sign-On authentication, specify the OracleAS Single Sign-On user name and password.

> **Note:** Oracle SES also supports crawling of Web sources that are protected by Oracle Access Manager (OAM) Single Sign-On mechanism. The following configurations must be done to use this feature:
>
> - The Web source that is protected by OAM Single Sign-On must be configured in Oracle SES with either the SSO login credentials or the HTTP authentication credentials.
>
> - The OAM server must be configured to send a basic authentication challenge in response to a request coming from Oracle SES. Refer to OAM documentation to know how to configure OAM to send basic authentication challenge for a specific User-Agent request.
>
> The following is the sequence of steps that get executed when the Oracle SES crawler tries to access the OAM protected Web source:
>
> 1. An HTTP request containing the appropriate User-Agent information is sent from the crawler to the Web source that is protected by OAM Single Sign-On.
>
> 2. OAM server responds with the status: 401 Basic Authentication.
>
> 3. The crawler reverts back to OAM server with the basic authentication credentials that are configured for the protected Web source in Oracle SES as either SSO login credentials or as HTTP authentication credentials.
>
> 4. The crawler gets access to the protected Web source after successful authentication.

Web Document Attributes

Oracle SES crawls and indexes these Web document attributes:

- Title

- Author

- Description

- Host

- Keywords

- Language

- LastModifiedDate

- Mimetype

- Subject: Mapped to "Description". If there is no description metatag in the HTML file, then it is ignored.

- Headline1: The highest H tag text; for example, "Annual Report" from `<H2>Annual Report</H2>` when there is no H1 tag in the page.

- Headline2: The second highest H tag text

- Reference Text: The anchor text from another Web page that points to this page.

You can define additional HTML metatags to map to a String attribute on the Home - Sources - Metatag Mapping page.

Setting Up Table Sources

A table source enables users to search a table or view in a relational database.

Database object names may be represented with a quoted identifier. A quoted identifier is case-sensitive and begins and ends with double quotation marks ("). If the database object is represented with a quoted identifier, then you must use the double quotation marks and the same case whenever you refer to that object.

When creating a table source in Oracle SES, if the table name is a quoted identifier, such as "1 (Table)", then in the **Table Name** field enter "1 (Table)", with the same case and double quotation marks. Similarly, if a primary key column or content column is named using a quoted identifier, then enter that name exactly as it appears in the database with double quotation marks.

> **See Also:** *Oracle Database SQL Reference* for more information about schema object names and qualifiers

The following procedures identify the basic steps for setting up a table source using the Oracle SES Administration GUI. For more information on each page, click **Help**.

To create a table source:

1. On the home page, select the **Sources** secondary tab to display the Sources page.

2. For Source Type, select **Table**.

3. Click **Create** to display the Create Table Source page.

4. Complete the following fields. Click **Help** for additional information.

   Database Information

   - **Source Name**: Name that you assign to this table source.

   - **Database Host Name**: Name or IP address of the host computer for the database.

   - **TCP Port Number**: Port number of Oracle Net Listener. The default port number for Oracle Database is 1521.

   - **SID**: System identifier or instance name of the database.

   - **User Name**: Database user name with SELECT access rights to the table or view.

   - **Password**: Password for **User Name**.

   - **Delete Passwords After Crawl**: Select to ensure that Oracle SES does not store the database schema password for this source longer than is needed to crawl the source.

   Table Information

   - **Schema**: Name of the database schema that owns the table or view.

   - **Table Name**: Name of the table or view.

   - **Primary Key Column**: Column or set of columns that uniquely identify each row in the table or view. For performance, the primary key must be indexed. Click **Locate Table** for a list of columns in the table.

   - **Content Column**: Column with searchable content for Oracle SES to crawl and index.

   - **Content Type**: Binary, plain text, or HTML.

5. Click **Create** or **Create & Customize**.

6. Follow the steps for crawling and indexing a source in "Getting Started Basics for the Oracle SES Administration GUI" on page 2-1.

To customize a table source:

1. When creating a table source, click **Create & Customize** on the Create Table Source page to display the Customize Table Source page.

   or

   After creating a source, click the **Edit** icon on the Home - Sources page.

2. Click the following subtabs and make the desired changes.

   ■ **Basic Settings**: Identifies the source name and schema password. Any other changes to the original settings require a new source definition.

   ■ **Table Column Mappings**: Maps columns to Oracle SES search attributes. See "Table Search Attributes" on page 6-6.

   ■ **Language**: Identifies the default language and different column languages.

   ■ **Display URL**: Specifies the URL that users see for security reasons instead of the actual URL.

   ■ **Authorization**: Configuration of an Access Control List or an authorization manager plug-in.

3. Click **Apply**.

Table Search Attributes

Table sources have no predefined attributes. The crawler collects attributes from columns defined during source creation. You must map the columns to the search attributes.

Setting Up File Sources

A file source enables users to search files on the local computer. The following procedures identify the basic steps for setting up a file source using the Oracle SES Administration GUI. For more information on each page, click **Help**.

To create a file source:

1. On the home page, select the **Sources** secondary tab to display the Sources page.

2. For Source Type, select **File**.

3. Click **Create** to display the Create File Source page.

4. Complete the following fields. Click **Help** for additional information.

   ■ **Source Name**: Name that you assign to this table source.

   ■ **Starting URL**: The URL of the top directory where the crawler begins. See "Tips for Using File Sources" on page 6-7.

5. Click **Create** or **Create & Customize**.

6. Follow the steps for crawling and indexing a source in "Getting Started Basics for the Oracle SES Administration GUI" on page 2-1.

To customize a file source:

1. When creating a file source, click **Create & Customize** on the Create File Source page to display the Customize File Source page.

    or

    After creating a source, click the **Edit** icon on the Home - Sources page.

2. Click the following subtabs and make the desired changes.

    ■   **Basic Settings**: Source name, language, and starting URL.

    ■   **URL Boundary Rules**: Contents of a URL that include or exclude a page from crawling.

    ■   **Document Types**: Common document and image types that you can include or exclude from crawling. By default, Oracle SES crawls HTML, Excel, PowerPoint, Word, PDF, and plain text.

    ■   **Display URL**: URL that users see for security reasons instead of the actual URL.

    ■   **Authorization**: Configuration of an Access Control List or an authorization manager plug-in.

    ■   **Attribute Mapping**: Maps document attributes to Oracle SES search attributes. See "File Document Attributes" on page 6-7.

    ■   **Crawling Parameters**: Crawling conditions, such as depth, language, HTTP cookies.

3. Click **Apply**.

File Document Attributes

Oracle SES crawls and searches various attributes. By default, Oracle SES maps these search attributes to common document attributes, such as AUTHOR, CREATOR, KEYWORD, and SUBJECT. You can enter and map additional document attributes.

Oracle SES crawls and indexes these document attributes:

■   Title

■   Author

■   Description

■   Host

■   Keywords

■   Language

■   LastModifiedDate

■   Mimetype

■   Subject

Tips for Using File Sources

This section contains the following topics:

■   Crawling File Sources with Non-ASCII Character Sets

■   Crawling File Sources with Symbolic Links

■   Crawling File URLs

- [Crawling File Sources from a Network Drive](#)
- [Crawling File Sources in a Clustered Middle Tier Environment](#)

Crawling File Sources with Non-ASCII Character Sets

For Oracle SES to crawl file sources that contain file names in non-ASCII format, Oracle SES server, Oracle ESS server, and WebLogic Server middle tier should be configured with the same locale settings that are used for the file names. Set the `LC_ALL` environment variable in the following files to the correct locale by editing these files in a text editor

Linux/UNIX Platform	Windows Platform
ses_home/bin/clexecutor.sh	*ses_home*\bin\clexecutor.cmd
ses_home/bin/searchctl	*ses_home*\bin\searchctl.bat
ses_home/ess/archives/ess-app/APP-INF/classes/META-INF/ess-config.xml	*ses_home*\ess\archives\ess-app\APP-INF\classes\META-INF\ess-config.xml
wls_home/common/bin/commEnv.sh	*wls_home*\common\bin\commEnv.bat

For example, for Korean environment, set `LC_ALL` to `ko_KR`. Restart the Oracle SES instance after setting the appropriate locale encoding for the `LC_ALL` environment variable in all the above files.

Crawling File Sources with Symbolic Links

When crawling file sources on UNIX, the crawler resolves any symbolic link to its true directory path and enforce the boundary rule on it. For example, suppose directory /tmp/A has two children, B and C, where C is a link to /tmp2/beta. The crawl has the following URLs:

- /tmp/A
- /tmp/A/B
- /tmp2/beta
- /tmp/A/C

If the inclusion rule is `/tmp/A`, then /tmp2/beta is excluded. The seed URL is treated as is.

> **Note:** When using Display URL feature to modify the file URL, note the following points:
>
> - File URL prefix should be a fully resolved path and it should not be a symbolic link.
> - When Display URL is specified, Inclusion/Exclusion rules get applied on the Display URL.

Crawling File URLs

For a plug-in to return file URLs to the crawler, the file URLs must be fully qualified. For example, `file://localhost/`.

If a file URL is to be used "as is", without going through Oracle SES to retrieve the file, then "file" in the Display URL Prefix should be upper case "FILE". For example, `FILE://localhost/...`. The starting URL is not case sensitive.

"As is" means that when a user clicks the search link of the document, the browser tries to use the specified file URL on the client computer to retrieve the file. Without that, Oracle SES uses this file URL on the server computer and sends the document through HTTP to the client computer.

Crawling File Sources from a Network Drive

If the files are crawled from a network drive, then the Oracle process should be started as a user who has access to the drive.

> **See Also:** "Required Tasks" on page 8-9 for instructions on how to change the user running the Oracle process.

Crawling File Sources in a Clustered Middle Tier Environment

When Oracle SES middle tier is deployed in a cluster, then for crawling a file source, all the nodes in the cluster should have access to the file source directory, that is, the file source directory must be shared across all the nodes in the cluster.

Setting Up E-Mail Sources

An e-mail source enables users to search e-mail messages on an IMAP server. The following procedures identify the basic steps for setting up an e-mail source using the Oracle SES Administration GUI. For more information on each page, click **Help**.

> **Note:** If Database and middle tier are present on different systems, then the e-mail source archive directory must be accessible from both these systems.

To create an e-mail source:

1. On the home page, select the **Sources** secondary tab to display the Sources page.

2. For Source Type, select **E-mail**.

3. Click **Create** to display the Create E-Mail Source page.

4. Complete the following fields. Click **Help** for additional information.

   - **Source Name**: Name that you assign to this table source.

   - **IMAP Server**: Address of the IMAP server, such as `mail.example.com`.

   - **Archive**: Directory where crawled messages are stored.

   - **Self Service**: **Enabled** to prompt users for their credentials, or **Disabled** to provide credentials for a single user on this page.

5. Click **Create**.

6. Follow the steps for crawling and indexing a source in "Getting Started Basics for the Oracle SES Administration GUI" on page 2-1.

E-Mail Attributes

Oracle SES crawls and searches these search attributes.

- Author

- Title

- Subject

- Language

- LastModifiedDate

Setting Up Mailing List Sources

A mailing list source enables users to search messages that were sent to a mailing list on an IMAP server.

The Oracle SES crawler is IMAP4 compliant. To crawl mailing list sources, you need an IMAP e-mail account. Oracle recommends that you create an e-mail account that is used solely for Oracle SES to crawl mailing list messages. The crawler is configured to crawl one IMAP account for all mailing list sources. Therefore, all mailing list messages to be crawled must be found in the Inbox of the e-mail account specified on this page. This e-mail account should be subscribed to all the mailing lists. New postings for all the mailing lists are sent to this single account and subsequently crawled.

Messages deleted from the global mailing list e-mail account are not removed from the Oracle SES index. The mailing list crawler deletes messages from the IMAP e-mail account as it crawls. The next time the IMAP account for mailing lists is crawled, the previous messages no longer exist. Any new messages in the account are added to the index and consequently deleted from the account. This keeps the global mailing list IMAP account clean. The Oracle SES index serves as a complete archive of all the mailing list messages.

> **Note:** For a Mailing List connector to work, you must create the directory `data/cache/mail` under Oracle home on the system where Database is installed.

The following procedures identify the basic steps for setting up a mailing list source using the Oracle SES Administration GUI. For more information on each page, click **Help**.

To create a mailing list source:

1. Enter the global mailing list settings:

   a. On the Global Settings page, choose **Mailing List Settings** under Sources to display the Global Mailing List Settings page.

   b. Complete the following fields. Click **Help** for additional information.

      User Name: IMAP e-mail account that is used to crawl the messages. This user must be on all of the mailing lists identified as a mailing list source.

      Password: Password for User Name.

      IMAP Server: Address of the IMAP server, such as `mail.example.com`.

   c. Click **Apply**.

2. On the home page, select the **Sources** secondary tab to display the Sources page.

3. For Source Type, select **Mailing List**.

4. Click **Create** to display the Create Mailing List Source page.

5. Complete the following fields. Click **Help** for additional information.

   - **Source Name**: Name that you assign to this table source.

   - **Mailing List**: Name of the mailing list to be searched, such as `news@example.com`.

6. Click **Create**.

7. Follow the steps for crawling and indexing in "Getting Started Basics for the Oracle SES Administration GUI" on page 2-1 for the mailing list schedule.

Mailing List Attributes

Oracle SES crawls and searches these search attributes:

- Author

- Title

- Subject

- Language

- LastModifiedDate

Setting Up OracleAS Portal Sources

An OracleAS Portal source enables users to search across multiple portal installations and repositories, such as Web pages, disk files, and pages on other OracleAS Portal instances. Oracle Secure Enterprise Search can securely crawl both public and private OracleAS Portal content.

> **Note:** Oracle SES 11.2.2.2 supports OracleAS Portal versions 10.1.2.0.2 and 10.1.4.2PS.

To create an OracleAS Portal source:

1. On the home page, select the **Sources** secondary tab to display the Sources page.

2. For Source Type, select **OracleAS Portal**.

3. Click **Create** to display the Create OracleAS Portal Source page.

4. Complete the following fields. Click **Help** for additional information.

   - **Source Name**: Name that you assign to this OracleAS Portal source.

   - **URL Base**: Base URL for OracleAS Portal.

   - **Page Groups**: List of page groups in OracleAS Portal retrieved when you click **Retrieve Page Groups**. Select the ones to crawl.

5. Click **Create & Customize**.

6. Select the **Authentication** tab.

7. Select **Enable OracleAS Single Sign-On Authentication** and enter your credentials.

8. Click **Apply**.

9. Follow the steps for crawling and indexing in "Getting Started Basics for the Oracle SES Administration GUI" on page 2-1 for the mailing list schedule.

Crawling a Folder or Page

The portal crawler can crawl a subtree under a specific folder or page instead of under an entire portal tree.

To set the boundary rule to crawl a specific folder or page:

1. On the home page, click the **Sources** secondary tab to display the Sources page.

2. Select a source and click **Edit** to display the Edit User-Defined Source page.

3. Click the **URL Boundary Rules** subtab.

4. Under **Inclusion Rules** for the URL, select the **starts with** rule and enter the value of the PORTAL_PATH for the folder or page.

    For example, to crawl only the P2 subtree of a portal tree, enter the path from the root to P2, such as /Proot/P1/P2.

OracleAS Portal Search Attributes

The crawler picks up key attributes offered by OracleAS Portal, as described in Table 6–1.

Table 6–1   OracleAS Portal Source Attributes

Attribute	Description
createdate	Date the document was created
creator	User name of the person who created the document
author	User-editable field so that they can specify a full name or whatever they want
page_path	Hierarchy path of the portal page/item in the portal tree (contains page titles)
portal_path	Hierarchy path of the portal page/item in the portal tree, used for browsing and boundary rules (contains page names) When searching OracleAS Portal 10.1.2, portal_path appears as upper case in the browse. When searching OracleAS Portal 10.1.4, portal_path appears in lowercase.
title	Title of the document
description	Brief description of the document
keywords	Keywords of the document
expiredate	Expiration date of the document
host	Portal host
infosource	Path of the Portal page in the browse hierarchy
language	Language of the portal page or item
lastmodifieddate	Last modified date of the document
mimetype	Usually 'text/html' for portal

Table 6–1 (Cont.) OracleAS Portal Source Attributes

Attribute	Description
perspectives	User-created markers that can be applied to pages or items, such as 'INTERNAL ONLY', 'REVIEWED', or 'DESIGN SPEC'. For example, a Portal containing recipes could have items representing recipes with perspectives such as 'Breakfast', 'Tea', 'Contains Nuts', 'Healthy' and one particular item could have several perspectives assigned to it.
wwsbr_name_	Internal name of the portal page or item
wwsbr_charset_	Character set of the portal page or item
wwsbr_category_	Category of the portal page or item
wwsbr_updatedate_	Date the last time the portal page or item was updated
wwsbr_updator_	Person who last updated the page or item
wwsbr_subtype_	Subtype of the portal page/item (for example, container)
wwsbr_itemtype_	Portal item type
wwsbr_mime_type_	Mimetype of the portal page or item
wwsbr_publishdate_	Date the portal page or item was published
wwsbr_version_number_	Version number of the portal item

Tips for Using OracleAS Portal Sources

- An OracleAS Portal source name cannot exceed 35 characters.

- URL boundary rules are not enforced for URL items. A URL item is the metadata that resides on the OracleAS Portal server. Oracle SES does not touch the display URL or the boundary rules for URL items.

- The `portal_path` attribute is used to compare boundary rules. Portal pages and items are organized in a tree structure. When a page is included or excluded, its entire subtree starting with that node is included or excluded.

- If OracleAS Portal user privileges change, the content the crawler collects might not be properly authorized. For example, in a Portal crawl, the user specified in the Home - Sources - Authentication page does not have privileges to see certain Portal pages. However, after privileges are granted to the user, on subsequent incremental crawls, the content still is not picked up by the crawler. Similarly, if privileges are revoked from the user, the content might still be picked up by the crawler.

  To be certain that Oracle SES has the correct set of documents, whenever a user's privileges change, update the crawler re-crawl policy to **Process All Documents** on the Home - Schedules - Edit Schedules page, and restart the crawl.

Setting Up Federated Sources

Secure federated search enables searching secure content across distributed Oracle SES instances. An end user is authenticated to the Oracle SES federation broker. Along with querying the secure content in its own index, the federation broker federates the query to each federation endpoint on behalf of the authenticated end user. This mechanism necessitates propagation of user identity between the Oracle SES instances. In building a secure federated search environment, an important consideration is the secure propagation of user identities between the Oracle SES instances. This section explains how Oracle SES performs secure federation.

To create a federated source:

1. On the home page, select the **Sources** secondary tab to display the Sources page.

2. For Source Type, select **Federated**.

3. Click **Create** to display the Create Federated Source page.

4. Complete the following fields. See "Federation Trusted Entities" on page 6-14 and click **Help** for additional information.

   - **Source Name**: Name that you assign to this federated source.

   - **Web Service URL**: The URL for the Web service.

   - **Remote Entity Name**: Name of the federation trusted entity on the federation endpoint.

   - **Remote Entity Password**: Password for Remote Entity Name.

   - **Search User Attribute**: Attribute used to authenticate users on the federation endpoint instance.

   - **Filter Rule**: Conditions for routing queries to this federated source. Filter rules can improve scalability. If no rule is defined, then the federation agent sends all queries to the federated source to perform the search.

5. Click **Create** or **Create & Customize**.

6. Follow the steps for crawling and indexing a source in "Getting Started Basics for the Oracle SES Administration GUI" on page 2-1.

To customize a federated source:

1. When creating a federated source, click **Create & Customize** on the Create File Source page to display the Customize File Source page.

   or

   After creating a source, click the **Edit** icon on the Home - Sources page.

2. Click the following subtabs and make the desired changes. See "Customizing Federated Sources" on page 6-17.

   - **Basic Settings**: Source name, Web Service URL, and so forth.

   - **Search Restrictions**: Controls whether the search is restricted, and if so, which source groups are searched.

   - **Attribute Retrieval**: Lists search attributes to retrieve at query time.

   - **Attribute Mapping**: Maps local and remote search attributes.

3. Click **Apply**.

   > **See Also:** "Configuring Secure Search with OracleAS Single Sign-On" on page 11-22

Federation Trusted Entities

When performing a secure search on a federation endpoint, the federation broker must pass the identity of the logged-in user to the federation endpoint. If the endpoint instance trusts the broker instance, then the broker instance can proxy as the end user. To establish this trust relationship, Oracle SES instances should exchange some secret. This secret is exchanged in the form of a *trusted entity*.

A trusted entity consists of two values: entity name and entity password. Each Oracle SES instance can have one or more trusted entities that it can use to participate in secure federated search. (A trusted entity is also referred to as a proxy user.)

An Oracle SES instance can connect to an identity management (IDM) system for managing users and groups. An IDM system can be an LDAP-compliant directory, such as Oracle Internet Directory or Active Directory.

Each trusted entity can be authenticated by either an IDM system or by the Oracle SES instance directly, independent of an IDM system. For authentication by an IDM system, check the box **Use Identity Plug-in for authentication** when creating a trusted entity. In this case, the entity password is not required. This is useful when there is a user configured in the IDM system that can be used for proxy authentication. Ensure that the entity name is the name of the user that exists in the IDM system and is going to be used as the proxy user.

For authentication of the proxy user by Oracle SES, deselect **Use Identity Plug-in for authentication** when creating a trusted entity. Then use any name and password pair to create a trusted entity.

Use **Authentication Attribute** to specify the format of the user credential that the Oracle SES federation endpoint expects for this particular trusted entity in proxy authentication. The identity plug-in registered on the federation endpoint should be able to map this user identity to the default authentication format used on the federation endpoint. This is useful when a federation broker cannot send user identity in the default authentication format used on the federation endpoint for proxy authentication, but the identity plug-in registered on the federation endpoint can map the value from the attribute in which it receives the user identity during proxy authentication to the default authentication format used on the federation endpoint.

To use a proxy entity, use the Web services API `proxyLogin` user name and password for the entity name and entity password. The identity plug-in can validate the password instead of storing it. When a request is sent for `proxyLogin`, Oracle SES calls the identity plug-in (which returns the call) to authenticate the entity. The `proxyLogin` must supply a valid trusted entity registered in the federation trusted entities.

User names are not case sensitive.

To perform secure federated search, both the broker and the endpoint instances involved in the federation must have identity plug-ins registered. The identity plug-ins may or may not talk to the same IDM system.

> **Note:** All user names should be unique across all Oracle SES instances. If not, then there should be a clear mapping for the users to make them unique across all IDMs involved in the secure federation.

Carefully specify the following parameters under the section **Secure Federated Search** when creating a federated source on the broker instance:

- **Remote Entity Name**: This is the name of the federation trusted entity on the federation endpoint. It is provided by the administrator of the endpoint instance.

- **Remote Entity Password**: This is the password of the federation trusted entity on the federation endpoint. It is provided by the administrator of the endpoint instance.

- **Search User Attribute**: This attribute identifies, and is used to authenticate, a user on the federation endpoint instance. This parameter is optional parameter, unless the broker and endpoint use different authentication attributes to identify end

users. For example, on the broker instance, an end user can be identified by user name; on the endpoint instance, the end user can be identified by e-mail address.

The identity plug-in registered on the broker instance should be able to map the user identity to this attribute based on the authentication attribute used during the registration of the identity plug-in. If this attribute is not specified during creation of the federation source, then the user identity on the broker instance is used to search on the endpoint instance.

> **Note:** If these parameters are not specified during the creation of the federated source, then the federated source is treated as a public source (that is, only public content is available to the search users).

- **Secure Oracle HTTP Server-Oracle SES channel**: Because any Oracle HTTP Server can potentially connect to the AJP13 port on the Oracle SES instances and masquerade as a specific person, either the channel between the Oracle HTTP Server and the Oracle SES instance must be SSL-enabled or the entire Oracle HTTP Server and Oracle SES instance computers must be protected by a fire wall.

> **Notes:**
>
> - In a secure federated search environment, the broker or the endpoint instance might or might not be using OracleAS Single Sign-on (SSO). However, the Web service URL of the endpoint should not be behind OracleAS Single Sign-on.
>
> - Oracle strongly recommends that you SSL-protect the channel between Oracle HTTP Server and Oracle SES for secure content. The endpoint instance should be SSL-enabled, or you should be able to access the Web service using HTTPS.

> **See Also:** "Tips for Using Federated Sources" on page 6-19

Example Creating a Federated Source

This section describes the steps for setting up a federated source that connects to Active Directory.

1. Activate the Active Directory identity plug-in on both the endpoint and broker instances. For example, on the Global Settings - Identity Management Setup page, enter the following:

   - **Parameter Name**: value

   - **Directory URL**: `ldap://ad.oracle.com:389`

   - **Directory account name**: administrator@ad.oracle.com

   - **Directory account password**: Password for **Directory account name**.

   - **Directory subscriber**: dc=ad,dc=oracle,dc=com

   - **Directory security protocol**: none

2. Create federation trusted entities on the endpoint instance. For example, login to Oracle SES on the endpoint instance, navigate to the Global Settings - Federation Trusted Entities page, and enter the following:

   - **Entity Name**: Entity name

- **Entity Password**: Password for **Entity Name**

3. Create a federated source on the broker side. For example, login to Oracle SES on the broker instance, navigate to the Home - Sources page, select the source type as Federated, and enter the following:

   - **Source Name**: Sourcename1

   - **Web Service URL**:
     `http://endpoint.cn.oracle.com:7777/search/query/OracleSearch`

   - **Remote Entity Name**: Entity name

   - **Remote Entity Password**: Password

4. To browse the federated source on broker side, create a source group and then add the federated source to the group.

Customizing Federated Sources

On the Home - Sources - Customize Federated Source page, you can change the source name, Web Service URL, remote entity name and password, and search user attribute.

This section describes the other ways you can customize a federated source:

- Route Queries to the Federated Source

- Set Search Restrictions

- Retrieve Attributes

- Map Attributes

Route Queries to the Federated Source

Enter a filter rule, which sets conditions for routing queries to the federated source, on the Home - Sources - Customize Federated Source page. Filter rules can improve scalability. If no rule is defined, then the federation agent sends all queries to the federated source to perform the search. The rules are applied only against the search query filter. They are not applied when an end user enters the attribute shortcut query.

Each rule has an attribute, a colon (:) and an expression. Rules can be based on end user properties, such as name or e-mail address, and on query information, such as document language, author, or document modified date. For example, an identity attribute could be `mail` or `dn`. A query attribute could be `author` or `lastmodifieddate`.

Multiple rules for the source are joined with the AND and OR operators. The attribute name and the operators are not case-sensitive. For example, the following rule defines that the federated source is for English documents and for users having an e-mail address starting with A in the identity management system:

```
(language:en ) AND (idm::mail:a.*)
```

The attribute can be Date, String, or Number type. For String attributes, the rule expression is regular expression. Oracle SES supports the regular expression syntax used in the JDK Pattern class (java.util.regex.Pattern). For Date and Number attributes, the expression contains the operator and value. The operators are =, >, >=, <, <=.

Filter Rule Examples The following rule defines that the federated source is for documents larger than 1 M:

```
content-length:>1000000
```

The following rule defines that the federated source is for documents published after 12/31/2006:

```
lastmodifieddate:> 12/31/2006
```

The following example defines that the federated source has only documents for the last week:

```
lastmodifieddate:> sysdate - 7
```

The following rule defines that the federated source is for the login name, which could be an attribute of the identity management repository:

```
username:test00.*
```

Set Search Restrictions

Restrict search to a specific list of source groups on the Home - Sources - Customize Federated Source - Search Restrictions page.

Available source groups from the federated source are retrieved when the page is loaded. When Source Group Restricted Search is selected, you can move the source groups between the **Not Searched** and **Searched** lists. When **Unrestricted Search** is selected, all source groups on the remote instance are searched.

The **Refresh Source Groups** button refreshes the available source groups from the remote instance. If a source group is no longer available, then it is marked **Not Available**. All newly available source groups after a refresh appear in the **Not Searched** list by default, and all existing source groups remain in the list they are presently in. If a remote source group is renamed, the old name is marked **Not Available** and the new name appears in the **Not Searched** list. Unavailable source groups persist while they remain in the Searched list.

If the federated source is unavailable, then the available source groups are loaded from local storage. A warning message then states that Oracle SES cannot retrieve the available source groups from the remote instance, indicating that the available source groups may be out of date.

> **Note:** A federated source can be restricted to only explicitly-created source groups on the remote Oracle SES instance. For example, a federated source cannot be restricted to the Miscellaneous group on the remote Oracle SES instance.

Retrieve Attributes

Identify which attributes to retrieve from the federated source on the Home - Sources - Customize Federated Source - Attribute Retrieval page.

Available attributes from the federated source are retrieved when the page is loaded. Move search attributes to retrieve between the **Not Retrieved** column and the **Retrieved** column. Attributes that are always retrieved by Oracle SES by default are in the **Retrieved** list and marked **Mandatory**. These attributes cannot be saved in the **Not Retrieved** list.

The **Refresh Attributes** button refreshes the available attributes from the remote instance. If an attribute is no longer available, then it is marked **(Not Available)**. All newly available attributes after a refresh appear in the **Not Retrieved** list by default, and all existing attributes remain in the list they are presently in. If a remote attribute is renamed, then the old attribute name is marked **Not Available**, and the new name

appears in the **Not Retrieved** list. Unavailable attributes persist while they remain on the Retrieved list or are used in an explicit attribute mapping.

If the federated source is unavailable, then the available attributes are loaded from local storage. A warning message then states that Oracle SES cannot retrieve the available attributes from the remote instance, so the available attributes may be out of date.

Map Attributes

Map local search attributes with federated search attributes on the Home - Sources - Customize Federated Source - Attribute Mapping page. For example, a local search attribute named Creator can be mapped to a remote attribute named Author. This is an explicit attribute mapping. Only one-to-one mappings between attributes of the same data type are supported.

> **Note:** For default Oracle SES search attributes, Oracle SES implicitly maps local attributes to remote attributes. For example, a remote attribute named Author is always mapped to local search attribute name Author. For all other attributes, explicit mappings must be created.

Local search attributes are the available attributes on the local instance, as defined on the Global Settings - Search Attributes page. Local search attributes that are used in a mapping cannot be deleted on the Global Settings - Search Attributes page. Initially, there are no mappings.

Remote search attributes are the available attributes on the federated source. This list is retrieved when the page is loaded. If a remote attribute is mapped to a local attribute but the remote attribute is no longer available, then the remote attribute is marked **(Not available)**. Only attribute mappings involving available remote attributes are used during queries.

Tips for Using Federated Sources

- The Oracle SES federator caches the federator configuration (that is, all federation-related parameters including federated sources). As a result, any change in the configuration takes effect within five minutes.

- If you entered proxy settings on the Global Settings - Proxy Settings page, then add the Web Services URL for the federated source as a proxy exception.

- If the federation endpoint instance is set to secure mode 3 (require login to search secure and public content), then all documents (ACL stamped or not) are secure. For secure federated search, create a trusted entity in the federation endpoint instance, then edit the federated source with the trusted entity user name and password.

- There can be consistency issues if you have configured a BIG-IP system as follows:

  - You have two Oracle SES instances configured identically (same crawls, same sources, and so on) behind a BIG-IP load balancer to act as a single logical Oracle SES instance.

  - You have two other Oracle SES instances configured identically along with Oracle HTTP Server and OracleAS Web Cache fronting each one and both servers behind BIG-IP. Each of these two instances federate to the logical

Oracle SES instance. Web Cache is clustered between these two nodes to act as a single logical Oracle SES instance called broker instance.

When a user performs a search on the broker Oracle SES instance and tries to access the documents in the result, document access may not be consistent each time. As a work-around, ensure that the load balancer sends all the requests in one user session to the exact same node each time.

Looping Among Federated Sources

A federation loop or cycle refers to a deployment in which multiple SES instances federate to each other. For example, if SES Instance A federates to SES Instance B, and SES Instance B federates back to instance A, then a federation cycle is in the deployment. Federation cycles can cause a flood of queries and high CPU load on the participating SES instances.

SES does not detect federation cycles, thus the Oracle SES administrator is responsible for avoiding them. You can explicitly remove them from the deployments or use source-group-restricted federation. The previous example can be fixed with a source-group restriction: the source groups on Instance B selected for federation on Instance A do not have any federated sources for Instance A, and the converse. See "Set Search Restrictions" on page 6-18.

Federated Search Characteristics

- Federated search can improve performance by distributing query processing on multiple computers. It can be an efficient way to scale up search service by adding a cluster of Oracle SES instances.

- The federated search quality depends on the network topology and the throughput of the entire federated Oracle SES environment.

Federated Search Limitations

- There is a size limit of 200KB for the cached documents existing on the federation endpoint to be displayed on the Oracle SES federation broker instance.

- For **infosource** browse, if the source hierarchies for both local and federated sources under one source group start with the same top level folder, then a sequence number is added to the folder name belonging to the federated source to distinguish the two hierarchies on the Browse page.

- For federated infosource browse, a federated source should be put under an explicitly created source group.

- On the Oracle SES federation broker, there is no direct access to documents on the federation endpoint through the display URL in the search result list for the following source types:

  – File (local files, not UNC)

  – Table

  – E-mail

  – Mailing list

  For these source types, only the cached version of documents is accessible.

7

Configuring Access to Content Management Sources

This chapter contains the following topics:

- Setting Up EMC Documentum Content Server Sources
- Setting Up Microsoft SharePoint Sources
- Setting Up Oracle Content Database Sources
- Setting Up Oracle Content Server Sources

Setting Up EMC Documentum Content Server Sources

Documentum data is stored in DocBases, which can contain cabinets and folders. A Documentum Content Server instance can have one or more DocBases crawled with an EMC Documentum Content Server source. The Documentum Content Server source navigates through the DocBases and the inline cabinets to crawl all the documents in Documentum Content Server. Oracle SES creates an index, stores the metadata, and accesses information in Oracle SES to provide search capabilities according to the end user permissions.

Oracle SES supports incremental crawling; that is, it crawls and indexes only those documents that have changed after the most recent crawling was scheduled. A document is re-crawled if either the content or metadata or the direct security access information of the document has changed. A document is also re-crawled if it is moved within Documentum Content Server and the end user has to access the same document with a different URL. Documents deleted from a DocBase are removed from the index during incremental crawling.

Important Notes for EMC Documentum Content Server Sources

The Documentum source in Oracle SES must use the administrator account of a DocBase for crawling and indexing documents of that DocBase.

Required Software

- Documentum Content Server DA (Documentum Administrator) *or* Documentum Content Server WebTop application must be installed and configured.
- Documentum Foundation Classes (DFC) must be installed on the server running Oracle SES.
- Currently supported Documentum version is 6.5.

Required Tasks

- Because EMC Documentum Content Server software is not included with Oracle SES, certain files must be copied manually into Oracle SES.

  The DFC installation asks for destination directory and user directory. For Windows, the default destination directory is `C:\Program Files\Documentum` and default user directory is `C:\Documentum`.

  For UNIX, you must create a DFC program root and a DFC user root. For example, DFC program root might be `user_home`/documentum_shared and DFC user root might be `user_home`/documentum.

- Copy the following files from the EMC Documentum Content Server to `ses_home`/search/lib/plugins/dcs/. These files may be stored in the `shared`, `dfc`, or `config` directories of the EMC Documentum Content Server.

  - `dctm.jar`

  - dfc.jar

  - dfcbase.jar

  - aspectjrt.jar

  - certjFIPS.jar

  - jsafeFIPS.jar

  - configservice-api.jar

  - dfc.properties

- Create a subdirectory in `ses_home`/search/lib/plugins/dcs/, for example `dcsothers`, and make a second copy of `dfc.properties` in it.

- Add the following to `DMCL.ini`:

  ```
  max_session_count = 20
  max_connection_per_session = 20
  ```

  In Windows, `DMCL.ini` is located in the `WINNT` folder. In Linux, `DMCL.ini` is available in the Documentum folder (DFC user root).

- In Windows 2003 Server, copy `dmcl40.dll` from `DFC_destination_directory`/shared/ to `ses_home`/bin. For UNIX platforms, copy `libdmcl40.so` from `DFC_destination_directory`/dfc/ to `ses_home`/lib32.

- The environment variables $DOCUMENTUM_SHARED (DFC Program root) and $DOCUMENTUM (DFC user directory) must be created before installing DFC on Linux. Also note that these variables must to be exported again, and Oracle SES must be restarted when the computer restarts. These variables can also be exported permanently in Linux.

  Use the following commands to export environmental variables in Linux:

  For DOCUMENTUM:

  ```
  export DOCUMENTUM=/home/sesuser/DOCUMENTUM
  ```

  For DOCUMENTUM_SHARED:

  ```
  export DOCUMENTUM_SHARED=/home/sesuser/DOCUMENTUM_SHARED
  ```

- Restart the middle tier.

  On Windows, restart the computer after installing DFC.

Known Issues

- In this release, search results cannot be viewed in Documentum desktop. The documents and folders can be viewed only using Documentum Administrator (DA) or Webtop applications.

- For the **Container name** parameter, a value of repository name alone might not work. Enter the value of *RepositoryName/CabinetName*. For example, *DocBaseName/CabinetName/FolderName/SubFolderName*.

- Incremental crawls do not recognize an ACL modification of access permissions from None to Browse and Browse to None. The DCSCHECKSUM attribute value is same for both settings.

Configuration for Documentum Content Server 6.5

For Windows, the JAR files can be taken from the application server directory where DA is deployed. For DFC installation on Linux, it is a prerequisite to create DFC program root and DFC user root. For example, the DFC program root can be *USER HOME*/DOCUMENTUM_SHARED and the DFC user root can be *USER HOME*/DOCUMENTUM. Table 7–1 lists the location of the JAR files in Windows and Linux.

Table 7–1   Location of the JAR Files

JAR File Name	Windows Location	Linux Location
dfc.jar	*Application server home directory*/*da deployment directory*/WEB-INF/lib	*DFC_destination_directory*
aspectjrt.jar	*Application server home directory*/*da deployment directory*/WEB-INF/lib	*DFC_destination_directory*/dfc
certjFIPS.jar	*Application server home directory*/*da deployment directory*/WEB-INF/lib	*DFC_destination_directory*/dfc
jsafeFIPS	*Application server home directory*/*da deployment directory*/WEB-INF/lib	*DFC_destination_directory*/dfc
dfc.properties	*Application server home directory*/*da deployment directory*/WEB-INF/classes	*DFC_user_directory*/config/
configservice-api.jar	*Application server home directory*/*da deployment directory*/WEB-INF/lib	*DFC_destination_directory*/dfc

To configure the crawler plug-in:

1. Create a new directory under *ses_home*/search/lib/plugin/dcs/. For example, dcsothers.

2. Copy dfc.properties to the folder created in the previous step (dcsothers) and to the main folder (dcs).

3. Copy dfc.jar, aspectjrt.jar, certjFIPS.jar, jsafeFIPS.jar, configservice-api.jar to the dcs folder in the following path *ses_home*/search/lib/plugin/dcs.

4. The environment variables $DOCUMENTUM_SHARED (DFC Program root) and $DOCUMENTUM (DFC user directory) must be created before installing DFC on Linux. Also note that the environment variables $DOCUMENTUM_SHARED,

$DOCUMENTUM, and $CLASSPATH must be exported again, and Oracle SES must be restarted when the computer restarts. These variables can also be exported permanently in Linux.

Use the following commands to export environmental variables in Linux:

For DOCUMENTUM:

```
export DOCUMENTUM=/home/sesuser/DOCUMENTUM
```

For DOCUMENTUM_SHARED:

```
export DOCUMENTUM_SHARED=/home/sesuser/DOCUMENTUM_SHARED
```

For CLASSPATH:

```
export CLASSPATH=$DOCUMENT_SHARED/dctm.jar:$DOCUMENTUM_SHARED/config
```

Setting Up Identity Management for EMC Documentum Content Server

Setting up identity management requires administration steps in both Oracle SES and EMC Documentum. It includes the following steps:

- Activating the Documentum Identity Plug-in
- Activating the OID Identity Plug-In
- Activating the AD Identity Plug-In
- Activating SunOne Identity Plug-In

Activating the Documentum Identity Plug-in

To activate the Documentum identity plug-in, perform the following steps:

1. Select **Documentum Identity Plug-in**.

2. Click **Activate**.

3. Enter a valid DocBase name.

4. Enter a valid user name and password.

5. Ensure that the environment variable DOCUMENTUM and DOCUMENTUM_SHARED are set correctly.

6. Click **Finish**.

Activating the OID Identity Plug-In

Before activating the OID Identity plug-in for validating the users in OID, Documentum Content Server should be synchronized with OID as an LDAP server. To synchronize them, you must import the users and groups from OID to Documentum:

1. Create an LDAP Configuration Object in Documentum Administrator (DA). To do this:

    a. Login to DA.

    b. Navigate to **Administration**, **User Management**, **LDAP**.

    c. In the File Menu, select **File**, **New**, **LDAP Configuration Object**.

    d. In the Name field, enter a name for LDAP Configuration Object.

    e. Select **dm_user** as the user subtype.

    f. Under Communication Mode, select **Regular**.

g. Under Import, select **Users** and **Groups**.

h. Select **Default Configuration Object** to use this configuration object in the server field.

i. Click **Next**.

j. In the Directory Type field, select **Oracle Internet Directory Server**.

k. In the Bind Type field, select **Bind by Searching for Distinguished Name**.

l. In the Binding Name field, provide the admin user name of OID. This is usually `cn=orcladmin`.

m. In the Binding Password field, provide the admin user password.

n. In the Host Name field, provide the OID host name.

o. Retain the default port number of OID (`389`).

p. In the Person Object Class field, provide the information of Base Person Object, typically the value is `inetOrgPerson`.

q. In the Person Search Base field, provide the person search base defined in OID. For example, `cn=Users, dc=us, dc=oracle, dc=com`.

r. In the Person Search Filter field, specify `cn=*`.

s. In the Group Object Class field, provide the Group Object. Typically the value is `groupOfUniqueNames`.

t. In the Group Search Filter field, specify `cn=*`.

u. Click **Next**.

v. The Attribute Map information is displayed. Click **Finish**.

2. Run the LDAP_Synchronization job. To do this:

a. Login to DA.

b. Navigate to **Administration**, **Job Management**, **Jobs**.

c. Open the job **dm_LDAPsynchronization**.

d. In the state field, select **Active**.

e. Select **Deactivate On Failure**.

f. In Designated Server, select the host name of Documentum Server.

g. Select **Run After Update**.

h. Click the **Schedule** tab.

i. In the **Start Date And Time** field, set the current date and time.

j. Select **Repeat time** from the Repeat list.

k. Set the Frequency field to any numeric value.

l. Select **End Date And Time** and specify how long the Synchronization job should run.

m. Click the `Method` tab.

n. Select **Pass Standard Argument**.

o. Click the **SysObject info** tab.

p. Click **OK**.

After synchronizing the Documentum Content Server with OID, you must activate the OID activity plug-in in Oracle SES. Perform the following steps:

1. Log in to the Oracle SES Administration GUI.

2. Click **Global Settings**.

3. Select **System**, **Identity Management Setup**.

4. Select **Oracle Internet Directory identity plug-in manager** and click **Activate**.

5. Select `nickname` from the **Authentication Attribute** list.

6. Provide the following values:

   - **Host name**: The host name of the computer where OID is running.

   - **Port**: The default LDAP port number, `389`.

   - **Use SSL**: `true` or `false` based on your preference.

   - **Realm**: The OID realm, for example, **dc=us.dc=oracle.dc=com**

   - **User name**: The OID admin username, for example, **cn=orcladmin**.

   - **Password**: User password

Activating the AD Identity Plug-In

Before activating AD Identity plug-in for validating the users in AD, Documentum Content Server must be synchronized with AD as an LDAP server. To synchronize them, you must import users and groups from AD to Documentum:

1. Create an LDAP Configuration Object in DA. To do this:

   a. Log in to DA.

   b. Navigate to **Administration**, **User Management**, **LDAP**.

   c. Select **File**, **New**, **LDAP Configuration Object**.

   d. Enter a name for ldap configuration object.

   e. Select **dm_user** as User Subtype.

   f. In the Communication Mode field, select **Regular**.

   g. In the Import field, select **Users** and **Groups**.

   h. Select **Default Configuration Object** in the server field, and click **Next**.

   i. Provide the following values:

      Directory Type: Select **Active Directory Server**.

      Bind Type: Select **Bind by Searching for Distinguished Name**

      Binding Name: Provide the admin user name of AD. It is normally **domainName/Administrator**.

      Binding Password: The password of the AD admin user.

      Host Name: AD host name.

      Port: Default port number of AD, `389`.

      Person Object Class: The Base Person Object, typically the value is `user`.

      Person Search Base: The person search base defined in AD, for example `cn=Users,dc=us, dc=oracle,dc=com`.

Person Search Filter: Enter `cn=*`.

Group Object Class: The group object. Typically the value is `group`.

Group Search Base: The group search base defined in AD. For example, `dc=us,dc=oracle,dc=com`.

Group Search Filter: Enter `cn=*`.

    j.   Click **Next**.

    k.   The Attribute Map information is displayed. Click **Finish**.

2. Run the LDAP_Synchronization job. To do this:

    a.   Login to DA.

    b.   Navigate to **Administration**, **Job Management**, **Jobs**.

    c.   Open the job **dm_LDAPsynchronization**.

    d.   In the state field, select **Active**.

    e.   Select **Deactivate On Failure**.

    f.   In Designated Server, select the host name of Documentum Server.

    g.   Select **Run After Update**.

    h.   Click the **Schedule** tab.

    i.   In the **Start Date And Time** field, set the current date and time.

    j.   Select **Repeat time** from the Repeat list.

    k.   Set the Frequency field to any numeric value.

    l.   Select **End Date And Time** and specify how long the Synchronization job should run.

    m. Click the `Method` tab.

    n.   Select **Pass Standard Argument**.

    o.   Click the **SysObject info** tab.

    p.   Click **OK**.

After the Documentum Content Server is synchronized with the AD, you must activate the identity for AD Identity plug-in. To perform this:

1. Log in to the Oracle SES Administration GUI.

2. Click **Global Settings**, and then select **System**, **Identity Management Setup**.

3. Select **Activity Directory Identity Plug-in Manager**, and click **Activate**.

4. Provide the following values:

- **Authentication Attribute**: Select `USER_NAME`.

- **Directory URL**: Provide the host name and the port number. For example, `ldap://ldapserverhost:port`.

- **Directory account name**: Provide the AD user name, for example `Administrator`.

- **Directory account password**: AD user password.

- **Directory subscriber**: Provide the directory subscriber (ldap base). For example, `dc=us.dc=oracle.dc=com`.

- ■ **Directory security protocol**: Specify either `none` or `portnumber`.

5. Click **Finish**.

Activating SunOne Identity Plug-In

Before activating SunOne Identity plug-in for validating the users in SunOne, you must synchronize Documentum Content Server with SunOne as an LDAP server. To synchronize them, you must import the users and groups from OID to Documentum:

1. Create an LDAP Configuration Object in DA:

- **a.** Log in to DA.
- **b.** Navigate to **Administration**, **User Management**, **LDAP**.
- **c.** Select **File**, **New**, **LDAP Configuration Object**.
- **d.** Enter a name for ldap configuration object.
- **e.** Select **dm_user** as User Subtype.
- **f.** In the Communication Mode field, select **Regular**.
- **g.** In the Import field, select **Users** and **Groups**.
- **h.** Select **Default Configuration Object** in the server field, and click **Next**.
- **i.** Provide the following values:

   Directory Type: Select **Netscape/iPlanet Directory Server**

   Bind Type: Select **Bind by Searching for Distinguished Name**

   Binding Name: Provide the admin user name of SunOne. It is normally **cn=Administrator**.

   Binding Password: The password of the SunOne admin user.

   Host Name: SunOne host name.

   Port: Enter the port number used for SunOne. The default port number of SunOne is `389`.

   Person Object Class: The Base Person Object, typically the value is `person`.

   Person Search Base: The person search base defined in SunOne, for example `cn=Users,dc=us, dc=oracle,dc=com`.

   Person Search Filter: Enter `cn=*`.

   Group Object Class: The group object. Typically the value is `groupOfUniqueNames`.

   Group Search Base: The group search base defined in AD. For example, `dc=us,dc=oracle,dc=com`.

   Group Search Filter: Enter `cn=*`.

- **j.** Click **Next**.
- **k.** The Attribute Map information is displayed. Click **Finish**.

2. Run the LDAP_Synchronization job:

- **a.** Login to DA.
- **b.** Navigate to **Administration**, **Job Management**, **Jobs**.
- **c.** Open the job **dm_LDAPsynchronization**.

d. In the state field, select **Active**.

e. Select **Deactivate On Failure**.

f. In Designated Server, select the host name of Documentum Server.

g. Select **Run After Update**.

h. Click the **Schedule** tab.

i. In the **Start Date And Time** field, set the current date and time.

j. Select **Repeat time** from the Repeat list.

k. Set the Frequency field to any numeric value.

l. Select **End Date And Time** and specify how long the Synchronization job should run.

m. Click the `Method` tab.

n. Select **Pass Standard Argument**.

o. Click the **SysObject info** tab.

p. Click **OK**.

After the Documentum Content Server is synchronized with SunOne, the identity is activated for SunOne Identity plug-in.

To activate the SunOne identity plug-in:

1. Log in to the Oracle SES Administration GUI.

2. Click **Global Settings**, and then select **System**, **Identity Management Setup**.

3. Select **Sun Java System Directory Server Manager**, and click **Activate**.

4. Provide the following values:

- **Authentication Attribute**: Select `USER_NAME`.

- **Directory URL**: Provide the host name and the port number. For example, `ldap://ldapserverhost:port`.

- **Directory account name**: Provide the Directory Server user name, for example `Administrator`.

- **Directory account password**: Directory Server user password.

- **Directory subscriber**: Provide the directory subscriber (ldap base). For example, `dc=us.dc=oracle.dc=com`.

- **Directory security protocol**: Specify either `none` or `portnumber`.

5. Click **Finish**.

Creating an EMC Documentum Content Server Source

Create an EMC Documentum Content Server source on the Home - Sources page. Select EMC Documentum Content Server from the Source Type list, and click **Create**. Enter values for the following parameters:

- **Container name**: The names of the containers to be crawled by Oracle SES. You can crawl an entire Documentum DocBase or a specific `repository/cabinet/folder`. The format is `DocBaseName/CabinetName/FolderName/SubFolderName`. Multiple comma-delimited container names can be entered. This parameter is

case-sensitive; hence, enter the exact same cabinet name as in the Documentum repository. Required

These are examples of container names:

- `DocBase1`: The entire DocBase1 is crawled.

- `DocBase2/Cabinet21`: Cabinet21 and its sub-folders within DocBase2 are crawled.

- `DocBase2/Cabinet21/Folder11`: Folder11 and its sub-folders are crawled.

- `DocBase1, DocBase2/Cabinet21/Folder11`: The entire DocBase1 and Folder 11 in DocBase2/Cabinet21 are crawled.

- **Attribute list**: The comma-delimited list of Documentum attributes along with their data types to be searchable. The format is `AttributeName:AttributeType, AttributeName:AttributeType`. Valid values are String, Number, and Date. See Table 7–2, " Documentum Data Type Mapping".

  While crawling a DocBase, an attribute is indexed only if both name and type match the configured name and type; otherwise, it is ignored. This is an optional parameter.

  For example, assume that you have the following Documentum attributes with the indicated data types

  - account name: String

  - account ID: Integer

  - creation date: Date

  To make these attributes searchable, enter this value for **Attribute list**:

  `Account Name:String, Account ID:Number, Creation Date:Date`

  The default searchable attributes for Documentum Content Server are Modified Date, Title, and Author.

  Multiple attributes with same name are not allowed, such as `Emp_ID:String` and `Emp_ID:Number`.

- **User name:** Enter the user name of a valid Documentum Content Server user. The user should be an administrator user or a user who has access to all cabinets, folders, and documents of the DocBases configured in the **Container name** parameter. The user should be able to retrieve content, metadata, and ACL from cabinets, folders, documents and other custom sub classes of all DocBases configured in **Container name** parameter. Required.

- **Password**: Password of the Documentum user. Required.

- **Crawl versions**: Indicate whether multiple versions of documents should be crawled, either `true` or `false`. The default value is `false`. Any other value is `false` and only the latest versions of a document are crawled. Optional.

- **Crawl folder attributes**: Indicate whether folder attributes must be crawled, either `true` or `false`. This is an optional parameter. The default value is `false`. Any other value is interpreted as `false`.

- **URL for viewing the documents**: A valid URL for Documentum WebTop or DA application used for viewing the Oracle SES search results. For example:

  `http://IP_address:port/da`

  or

```
http://IP_address:port/webtop
```

- **Authentication Attribute**: This parameter is used to set ACLs. This parameter lets you set multiple LDAP servers. If Oracle SES and Documentum Content Server are synchronized with Active Directory, then enter the value USER_NAME. If Oracle Internet Directory is used, then enter nickname.

Table 7–2   Documentum Data Type Mapping

Sr. No	Documentum Data Type	Oracle SES Data Type
1	Boolean	Number
2	Integer	Number
3	String	String
4	ID	String
5	Time or Date	Date
6	Double	Number

Setting Up Microsoft SharePoint Sources

The SharePoint Crawler connector enables Oracle SES to provide secure search over SharePoint Portal Server 2003 (SPPS 2003), Microsoft Office SharePoint Server 2007 (MOSS 2007), and Microsoft SharePoint Server 2010 (SPS 2010). The connector extends the searching capabilities of Oracle SES and enables it to search into an external repository. Oracle SES can crawl through the documents, items, and related metadata in SharePoint repositories and provide secure, full-text search. The connector also provides metadata search and browse functionality, which allows a search to be done against a specific subfolder in the hierarchy.

In SharePoint, data is stored in different libraries such as the Document Library, Picture Library, Lists, Discussion Boards, and so on. A SharePoint instance can have one or more sites and sub-sites that the SharePoint Crawler connector can crawl after you set up the appropriate configuration parameters in the Oracle SES Administration GUI. The SharePoint Crawler connector navigates through the Libraries and Lists to crawl all the documents and items from a SharePoint repository. It creates an index, stores the metadata, and accesses information in Oracle SES to provide search capabilities according to the end user permissions.

The SharePoint Crawler connector supports incremental crawling, which means that it crawls and indexes only those documents that have changed after the most recent crawl. A document is re-crawled if the content, metadata, or direct security access information of the document has changed since the previous crawl. Documents deleted from a Library are removed from the index during incremental crawling.

Important Notes About SharePoint Sources

- The supported versions of SharePoint Server are:
  - 2003 or 2.0 for SharePoint Portal Server (SPPS)
  - 2007 or 3.0 for Microsoft Office SharePoint Server (MOSS)
  - 2010 for Microsoft SharePoint Server (SPS)
- When the **Crawl Security Settings** parameter is set to either NORMAL or STRICT, the SharePoint Crawler for the Container must use the SharePoint administrator account for crawling and indexing documents.

- When the **Crawl Security Settings** parameter is set to `RELAX`, any user that has at least Visitor (Read) permissions can be identified in the SharePoint source for crawling and indexing documents.

- SharePoint Container names in Oracle SES should not contain any special characters. Enter a backslash (\) before a slash or a comma. Otherwise, the crawler does not recognize the Container.

Known Limitations of SharePoint Connector

- Passwords entered through the Oracle SES Administration GUI are case insensitive.

- Storing more than 200 files in a single folder may result in degraded performance and increased crawling time.

- An administrator must own the SharePoint Server site with the documents to be crawled. The crawler does not have sufficient access rights to crawl the documents if it uses the identity of a non-administrative user.

  To grant administrative rights to a SharePoint user:

  1. Open the SharePoint Site UI and select **Site Settings**.

  2. Select **Users and Permissions**, then **Site Collection Administrators** to display the Site Collection Administrators page.

  3. Enter the user name of the SharePoint Server site in the Site Collection Administrators field.

  4. Click **OK**.

- If the Crawler Security Settings parameter is set to `RELAX`, then the user ID specified in the User Name parameter does not require administrative privileges. Visitor (Read) permissions on the site are sufficient. However, Read must have Browse Directories permissions to access any sub-sites. Otherwise, the sub-sites are not crawled.

  To add Browse Directories permissions for MOSS 2007/SPS 2010:

  1. Open People and Groups - Site Permissions.

  2. Under Settings - Permission Levels, select **READ**.

  3. Under Site Permissions, select **Browse Directories**.

  4. Click **Submit**.

  To add Browse Directories permissions for SPPS 2003:

  1. Open the Created subarea and select **Manage Security**.

  2. Select the user and edit permissions.

  3. Select READ.

  4. Click **Advanced Permissions**.

  5. Under Advanced Permissions, select **Browse Directories**.

  6. Click **OK**.

- SharePoint does not allow users without administrative privileges to browse user profiles.

  If the user ID specified in the User Name parameter does not have administrative privileges, then this user needs permission to manage profiles.

To grant permission to manage profiles:

1. Open SharePoint Central Administration 3.02.

2. Click **Shared Services Administration - SharedServices1**.

3. Under **User Profile** and **My Sites**, select **Personalization Service Permissions**.

4. Add user *user1* and select permissions **Manage user profiles**.

5. Save and submit the user.

User profiles are crawled if the user has specified the root site in the Site/Sub-Site URL parameter of the source configuration.

Known Issues for SharePoint Connector

- Versions of list items whose object type is Folder are not crawled and indexed.

- Site Collection Administrator users are not able to see documents if they are not listed among the document permission users.

- `Unable to type cast null` message is not error. This information is provided when the crawler tries to crawl attachments that are not supported for a particular entity.

- `Principal user_name cannot be validated` error is returned when the crawler obtains a user name from the SharePoint repository that is not present in the Active Directory.

- Performance of the SharePoint connector can be impacted when the Crawl Versions attribute is set to true.

Supported Platforms

The following platforms are supported by the SharePoint Crawler connector:

- Red Hat Linux 4

- Windows 2003 Server Standard Edition and later with the latest Service Pack

Creating a Microsoft SharePoint Source

Create a source for the newly-created user-defined source type on the Home - Sources page. Enter a source name. Provide values for the configuration parameters described in the following list. Also see Table 7–3, " Supported Values for SharePoint Source Parameters".

- **SharePoint Version**: Version of the SharePoint server (SPPS 2003 or MOSS 2007 or SPS 2010) to crawl. (Required)

- **Container name**: Contains the names of the containers to be crawled by Oracle SES. You can specify multiple container names as a comma-delimited list. (Required)

  You can crawl an entire area or site or a specific folder. The format for specifying a container folder is *AreaName/LibraryName/FolderName/SubFolderName*.

  To crawl all documents in the Area or Library, the format is *AreaName* or *AreaName/LibraryName*.

  To index the entire SharePoint portal, enter a slash (/).

  To crawl all sites, enter `sites`.

Examples for SharePoint Portal Server:

– Container name: *AreaName*

The entire Area is crawled.

– Container name: *AreaName/LibraryName/Folder21*

Folder21 and its subfolders within *LibraryName* are crawled.

– Container name: *LibraryName*

All documents inside the Library and its subfolders are crawled.

Examples for MOSS 2007/SPS 2010:

– Container name: *LibraryName/Folder21*

Folder21 and its sub-folders within *LibraryName* are crawled.

– Container name: *LibraryName*

All documents inside the Library and its subfolders are crawled.

The path for the container cannot contain any special characters. Enter a backslash (\) before a slash or a comma.

- **Attribute list**: A comma-delimited list of attributes, as described in Table 7–4. The format for an attribute list is *AttributeName*, *AttributeName*. Multiple attributes with same name are not allowed, such as Emp_ID, Emp_ID.

  In MOSS 2007/SPS 2010, all attributes viewable from the UI are indexed by default. List all custom attributes to index, using the names displayed in the user interface.

  In SPPS 2003, the Title, LastModifiedDate, and Author attributes are indexed by default. List any other attributes to index, using the names displayed in the UI.

  If you update the attribute list from the administrator parameters, then perform a forced recrawl to delete the indexes of the old attribute list and to create indexes for the new attribute list.

- **Domain name**: The domain name of the user that is used to crawl the SharePoint site. For example, if you intend to use the OracleDomain\Administrator user for crawling, then enter OracleDomain for this parameter. Do not include .com or .in or any other suffix in the name. (Required)

- **User name**: Specifies the user name of a valid SPPS 2003/MOSS 2007/SPS 2010 user. Do not include the domain name for this user. For example, for OracleDomain\Administrator, enter Administrator. (Required)

- **Password**: Specifies the password of the SharePoint user specified in User name. (Required)

- **Authentication attribute**: Format of the user and group identity stored in the ACL of SharePoint objects. This format must be an authentication attribute of the Oracle SES active identity plug-in, such as USER_NAME for an Active Directory identity plug-in. Otherwise, the ACL validation fails during indexing. (Required and case sensitive)

  For example, this value is USER_NAME for the Microsoft Active Directory identity plug-in.

- **Crawl URL or SPS Site/Sub-Site URL**: The URL of the Site or Sub-site of the Microsoft SharePoint source, which is used for viewing the search results. (Required)

This URL has the form `http://HostName:PortNumber` or `http://HostName:PortNumber/SubSiteName`.

- **Crawl Security Settings**: Sets security on documents for indexing. (Required)

  This setting can be one of the following:

  - `NORMAL`: The regular crawl uses site-level access control lists (ACLs) but not document-level ACLs.

  - `RELAX`: When the SharePoint Site Administrator user information is not available and the SharePoint user has visitor (or read) permissions on the site, this user is not able to crawl subsites under the main site. This mode is intended for exposing public documents temporarily and quickly to search. The SES administrator must be careful not to expose documents to other users inadvertently. See the work-around for this in "Known Limitations of SharePoint Connector" on page 7-12.

  - `STRICT`: Captures even document-level security. This mode requires that an additional Web Service agent, Oracle MOSS Web Service, be installed on the MOSS 2007/SPS 2010 server. See "Deploying the Web Service on MOSS 2007/SPS 2010" on page 7-17.

- **Simple Include**: Only include URLs having at least one word mentioned in this parameter. Separate the words with commas.

- **Simple Exclude**: Exclude all URLs having one or more word(s) mentioned in this parameter. Separate the words with commas.

- **Regular Expression Include**: Include all URLs that match the expression provided in this parameter.

- **Regular Expression Exclude**: Exclude all URLs that match the expression provided in this parameter.

- **Crawl versions**: Controls whether multiple versions of documents are crawled. Valid values are `true` and `false`. Any other value is interpreted as `false`. The default value is `false`, so only the latest version is crawled. (Optional)

- **Crawl folder attributes**: Controls whether folder attributes are crawled. The default value is `false`. Valid values are `true` or `false`, and any other value is interpreted as `false`. (Optional)

- **Crawl attachments**: This parameter indicates whether attachments should be crawled. The default value is `false`. Valid values are `true` or `false`, and any other value is interpreted as `false`. (Optional)

- **LDAP URL**: URL of the LDAP server, such as `ldap://IP:port`, where the default port number is 389.

- **LDAP Search Base**: LDAP Search Base, such as, `DC=abc`, `DC=com`. When the value of **Authentication Attribute** is DN, specify the LDAP URL and the LDAP search base of the LDAP server configured in the identity plug-in. Otherwise, leave these parameters blank.

- **Fast Incremental Crawl**: Controls whether the fast incremental crawl should be used while crawling. Enter `true` to enable the fast incremental crawl. Setting to `true` fetches newly added and modified documents, without removing the documents from the Oracle SES index that were deleted from the Sharepoint repository. Setting to `false` fetches newly added and modified documents, and removes the documents from the Oracle SES index that were deleted from the Sharepoint repository.

Table 7–3 summarizes the supported values for the configuration parameters of the SharePoint Crawler connector.

Table 7–3   Supported Values for SharePoint Source Parameters

Parameter Name	SharePoint Portal Server	MOSS 2007/SPS 2010
SharePoint Version	2003, 2.0	2007, 3.0 / 2010
Container name	(/) for full site, Library Name, List Name, Area Name	(/) for full site, Library Name, List Name
Attribute list	`AttributeName1, AttributeName2`	`AttributeName1, AttributeName2`
Domain Name	Domain name of the user	Domain name of the user
User name	Valid administrator user for SharePoint Portal server	Valid administrator user for MOSS 2007/SPS 2010
Password	Password for the user	Password for the user
Authentication attributes	`USER_NAME`	`USER_NAME`
SPC Site/Sub-Site URL	IP address or host name with port on which SharePoint Portal Server is installed	IP address or host name with port on which MOSS 2007/SPS 2010 is installed
Crawl Security Settings	`NORMAL`, `RELAX`	`NORMAL`, `RELAX`, `STRICT`
Simple Include	Part of URL	Part of URL
Simple Exclude	Part of URL	Part of URL
Regular Expression Include	All URLs that match the expression	All URLs that match the expression
Regular Expression Exclude	All URLs that match the expression	All URLs that match the expression
Crawl versions	`true` or `false`	`true` or `false`
Crawl folder attachments	`true` or `false`	`true` or `false`
Crawl attachments	`true` or `false`	`true` or `false`
LDAP URL	URL of the LDAP server	URL of the LDAP server
LDAP Search Base	LDAP Search Base	LDAP Search Base
Fast Incremental Crawl	`true` or `false`	`true` or `false`

Table 7–4   Attributes for List Items and Versions Crawled for SharePoint Source

List Item Type	Attributes
Document Library	Title, Author, Created, Modified
Picture Library	Title, ImageSize, ImageCreateDate, Description, Keywords
Form Library	Title, Author, Created, Modified
Translation Library	Title, Name, Language, Base Document Version, Translation Status, Created
Data Connection Library	Connection Type, Description, Keywords, Title, UDC Purpose, Created
Slide Library	Name, Presentation, Description, Created

Table 7–4 (Cont.) Attributes for List Items and Versions Crawled for SharePoint Source

List Item Type	Attributes
Report Library	Name, Title, Author, Created, Report Category, Report Status
Dash Board	Name, Title, Author, Created
Wiki Page Library	Title, Author, Created, Modified
Announcements	Title, Body, Editor, Modified, Author, Created
Contacts	Company, WorkCity, Created, Email, Comments, Title, Editor, HomePhone, JobTitle, Modified, WorkZip, WorkPhone, WorkState, FirstName, Author, FullName, WorkCountry, CellPhone, WorkFax, WorkAddress
Links	Comments, Editor, Modified, Author, URL, Created
Discussion Reply	Body, Created, DiscussionTitle, Editor, Modified, Author
Calendar	EventType, Title, EventDate, Duration, Editor, WorkspaceLink, Modified, EndDate, Description, fRecurrence, Author, fAllDayEvent, Created
Task	Title, StartDate, Body, Status, Editor, Priority, AssignedTo, DueDate,Modified, Author, PercentComplete, Created
Project Task	Title, StartDate, Body, Status, Editor, Priority, AssignedTo, DueDate,Modified, Author, PercentComplete, Created
Issue Tracking	Category, LinkIssueIDNoMenu, RelatedIssues, IssueID, Priority, DueData, Comment, V3Comments, IsCurrent, Created, Title, Status, Editor, AssignedTo, Modified, Author
Custom List	Title, Editor, Modified, Author, Created
Languages and Translators	Language_x0020_From,Language_x0020_ To,Modified,Author,Translator,Created, Editor
KPI List	Title, PercentExpression, Editor, ViewGuid, Modified, Value, AutoUpdate, KpiComments, Author, Goal, ValueExpression, Warning, KpiDescription, DataSource, LowerValuesAreBetter, Created
Asset Library	Title, Description, Author, CreatedDate
Publishing Page	Title, Name, Description, CreatedDate, Modified
Site Pages	Title, Name, Description, CreatedDate, Modified

Deploying the Web Service on MOSS 2007/SPS 2010

For MOSS 2007/SPS 2010, if the **Crawl Security Settings** parameter is set to `STRICT`, then you must install an extra web service, Oracle MOSS Web Service. The following installation and deinstallation files are provided by the OracleMOSSService installer at *ses_home*/search/lib/plugins/sps/WebService.zip:

- OracleMossService.wsp

- install.cmd

- de-install.cmd

To install or deinstall the Oracle MOSS Web Service:

1. Click install.cmd to install, or click de-install.cmd to deinstall.

2. Verify that the STSADM.exe file is in the following location: *Drive*:\Program Files\Common Files\Microsoft Shared\web server extensions\12\BIN.

If STSADM.exe is not in that folder, specify the correct path when the installer prompts for it.

3. Press any key to continue.

Setting Up Oracle Content Database Sources

Documents in Oracle Content Database are organized into **folders**. Oracle SES navigates the folder hierarchy to crawl all documents in Oracle Content Database. It creates an index, stores the metadata, and accesses information in Oracle SES to provide search according to the end users' permissions.

The metadata crawled includes `folder_url` (URL of the folder containing the document) and `folder_path` (path of the folder containing the document). These let you show the direct folder path and direct folder URL for each document hit.

Oracle SES supports incremental crawling; that is, it only crawls and indexes documents that have changed since the last crawling. A document is re-crawled if either the content or the direct security access information of the document changes. A document is also re-crawled if it is moved within Oracle Content Database and the end user has to access the same document with a different URL. Deleted documents are removed from the index during incremental crawling.

Important Notes for Oracle Content Database Sources

This book uses the product name Oracle Content Database to mean *both* Oracle Content Database *and* Oracle Content Services. Oracle Content Database sources are certified with Oracle Content Database release 10.2 and release 10.1.3 and Oracle Content Services release 10.1.2.3.

Known Issues:

- The administrator account used by the Oracle Content Database source must have the **ContentAdministrator** role on the site that is being crawled and indexed. Also, end users searching documents in Oracle Content Database must have the **GetContent** and **GetMetadata** permissions.

- By default, Oracle Content Database has a limit of three concurrent requests (simultaneous operations) for each user. However, Oracle SES has a default of five concurrent crawler threads. When crawling Oracle Content Database, only three of the five threads can successfully crawl, which causes the crawl to fail.

  Workaround: For an Oracle Content Database source, change the **Number of Crawler Threads** on the Home - Sources - Crawling Parameters page to a value of 3 or fewer.

  Or, modify the Oracle Collaboration Suite configuration in Oracle Enterprise Manager to allow more than three concurrent requests. For example:

  1. Access the Enterprise Manager page for the Collaboration Suite Midtier. For example: `http://example.domain:1156/`.

  2. Click the Oracle Collaboration Suite midtier standalone instance name. For example: `ocsapps.example.domain`.

  3. In the **System Components** table, click **Content**.

  4. From **Administration**, click **Node Configurations**.

  5. In the **Node Configurations** table, click **HTTP_Node**. For example: ocsapps.computer.domain_HTTP_Node.

6. On **Properties**, change the value for **Maximum Concurrent Requests Per User**. Enter a value larger than or equal to the number of crawling threads used by Oracle SES. This value is listed on the Global Settings - Crawler Configuration page.

Setting Up Identity Management for Oracle Content Database Sources

The Oracle SES instance and the Oracle Content Database instance must be connected to the same or mirrored Oracle Internet Directory system or other LDAP server.

To set up a secure Oracle Content Database source:

1. Read "Known Issues:" on page 7-18 and confirm that the number of crawler threads does not exceed the available concurrent connection settings for each user in Oracle Content Database.

2. Activate the Oracle Internet Directory identity plug-in for the Oracle Content Database instance on the Global Settings - Identity Management Setup page in Oracle SES.

3. For Oracle Content Database 10.1.2.3 and 10.2.0.4, use the following LDIF file to create an *application entity* for the plug-in. (An application entity is a data structure within LDAP used to represent and keep track of software applications accessing the directory with an LDAP client.)

   ```
   ses_home/bin/ldapmodify -h oidHost -p OIDPortNumber -D "cn=orcladmin" -w
   password -f  ses_home/search/config/ldif/csPlugin.ldif
   ```

   This defines the entity that is used for the connector:
   `orclApplicationCommonName=ocsCsPlugin,cn=ifs,cn=products,cn=oraclecontext`. The entity has the password `welcome1`.

Creating an Oracle Content Database JDBC Source

The Content Database JDBC connector is an alternative to the Content Database connector provided in Oracle SES Release 10.1. The JDBC connector greatly improves the performance of incremental crawls. If the elapsed time of an incremental crawl is an important consideration in your deployment of Oracle SES, then use the JDBC connector.

Oracle SES crawler supports crawling from Oracle Content Database 10.1.2.0.4 or later. See the readme file for Oracle Content Database 10.2.1.0.4 patchset for details on configuring high volume full and incremental crawls in Oracle Content Database.

You may need to grant the SES user access to a Oracle Content Database object. Use this command:

```
GRANT SELECT ON ODMC_ALERT_SEQ TO sesuser
```

where `sesuser` is the SES user.

For example,

```
GRANT SELECT ON ODMC_ALERT_SEQ TO searchsys
```

> **Note:** The JDBC connector requires installation of a patch to Oracle Content Database. If the patch is not available for your version of Content Database, then use the older connector as described in "Creating an Oracle Content Database Source" on page 7-21.

To create an Oracle Content Database JDBC source:

1. Open the Oracle SES Administration GUI.

2. On the home page, select the Sources secondary tab.

3. For Source Type, select **Oracle Content Database (JDBC)**, then click **Create** to display Step 1 Parameters.

4. Enter the source name and the values for the parameters described in Table 7–5.

5. Click **Next** to display Step 2 Authorization.

6. Enter the settings described in Table 7–6.

7. Click **Create** or **Create and Customize** to create the source.

Table 7–5   Oracle Content Database JDBC Source Parameters (Step 1)

Parameter	Value
Database Connection String	JDBC connection string to Oracle Content Database in the form `jdbc:oracle:thin@server:port:sid`. For example, `jdbc:oracle:thin@example.com:1521:rel11g`
Content DB System User	SYSTEM user for Content Database.
Alert Table Name	Name of the Alert table for Content Database, which typically has the form `ODMC_ALERT_name`.
Database User ID for Crawl	Valid user ID for the Content DB database.
Database Password for Crawl	Password associated with the user ID for crawling.
Document Count	Maximum number of documents to be crawled.
URL Prefix	URL to Oracle Content Database in the form `HTTP://hostname:port/CONTENT`. For example, `HTTP://example.com:7778/CONTENT`.
Document Access (DAV) User ID	Valid Content Database user ID for using WebDAV to access documents.
Document Access (DAV) Password	Password associated with the DAV user ID.
Starting Path for Crawl	Full path where the crawl starts. Enter `/` to crawl the entire Content Database hierarchy.

Table 7–6   Oracle Content Database JDBC Authorization Parameters (Step 2)

Parameter	Value
Authorization Database JDBC Connection String	JDBC connection string to Oracle Content Database in the form `jdbc:oracle:thin@server:port:sid`. For example, `jdbc:oracle:thin@example.com:1521:rel11g`
Content DB System User	System user for Content Database, such as `CONTENT` or `IFS_SYS`.
Database User ID	User ID to connect to the database.
Database Password	Password associated with the database user ID.
Use the Run-Time Result Filter	Controls use of a final security check: `TRUE`: Performs a final security check on each row in the result set. `FALSE`: Does not do a final check. (Default)

Table 7–6   (Cont.) Oracle Content Database JDBC Authorization Parameters (Step 2)

Parameter	Value
Authorization User ID Format	Format of user ID in the authorization query. Enter a supported authentication attributes of the active ID plugin, such as `nickname`.

Creating an Oracle Content Database Source

If Oracle Content Database release 10.2 or Oracle Content Services release 10.1.2 is used, then the **Entity name** and **Entity password** parameters are required, the last six parameters related with keystore are not required, and the crawler plug-in uses service to service (S2S) authentication to connect to Oracle Content Database.

If Oracle Content Database release 10.1.3 is used, then the last six parameters in the following table are required, the **Entity name** and **Entity password** are not required, and Oracle SES uses Web services authentication to connect to Oracle Content Database. See "Required Tasks for Oracle Content Database Release 10.1.3" on page 7-22.

Create an Oracle Content Database source on the Home - Sources page. Select **Oracle Content Database** from the Source Type list, and click **Create**.

Enter values for the parameters listed in Table 7–7.

Table 7–7   Oracle Content Database Source Parameters

Parameter	Value
Oracle Content Database URL	`http://`*host name:port*`/content`
Starting paths	`/`
Depth	-1
Oracle Content Database admin user	`orcladmin`
Entity name	`orclApplicationCommonName=ocsCsPlugin,cn=ifs,cn=products,cn=oraclecontext`
Entity password	welcome1
Crawl only	`false`
Use e-mail for authorization	`false`
Oracle Content Database Version	For example, 10.1.3.2.0
SES keystore location	For example, /scratch/ocs/cdb/cdb-ses/keystore/sesClientKeystore.jks
SES keystore type	jks
SES keystore password	*******
SES private key alias	client
SES private key password	*******
CDB Server public key alias	server

Table 7–8   Oracle Content Database Authorization Manager Plug-in Parameters

Parameter	Value
Oracle Content Database URL	http://*host name*:*port*/content
Oracle Content Database admin user	orcladmin
Entity name	orclApplicationCommonName=ocsCsPlugin,cn=ifs,cn=products,cn=oraclecontext
Entity password	welcome1
Use e-mail for authorization	false
Use result filter for authorization	false You can use a real-time result filter (query-time authorization) to ensure that the user has access to each result document. Set this parameter to true to remove documents that the user has lost access to since the last crawl.
Oracle Content Database Version	For example, 10.1.3.2.0
SES keystore location	For example, /scratch/ocs/cdb/cdb-ses/keystore/sesClientKeystore.jks
SES keystore type	jks
SES keystore password	********
SES private key alias	client
SES private key password	*******
CDB Server public key alias	server

Required Tasks for Oracle Content Database Release 10.1.3

This section describes the required steps for Web services authentication when using Oracle Content Database release 10.1.3. This procedure uses the JDK keytool to create the keys.

> **Note:** In the following steps, content_db_oracle_home denotes the Oracle home of the Oracle Content Database server and ses_oracle_home denotes the Oracle home of the Oracle SES application.

> **See Also:** "Setting Up a Server Keystore for WS-Security" in the *Oracle Fusion Middleware Administrator's Guide for Oracle Universal Online Archive*

1. Configure a server keystore at the Oracle Content Database middle tier if the keystore is not set up yet.

   The file *content_db_oracle_home*/oc4j/j2ee/home/config/oc4j.properties defines the keystore type and the keystore properties file location. If you use a different file name for the keystore, then edit the file on the following entry:

   ```
   oracle.ifs.security.KeyStoreLocation=/home/oracle/product/10.1.3.2.0/OracleAS_1/content/settings/server-keystore.jks
   ```

   a. Change to the settings directory:

      ```
      cd content_db_oracle_home/content/settings
      ```

b. Create the Oracle Content Database server keystore with the following keytool command, substituting a secure password for *password*.

```
content_db_oracle_home/jdk/bin/keytool -genkey -keyalg RSA -validity 5000
-alias server -keystore server-keystore.jks -dname "cn=server" -keypass
password -storepass password
```

To list the keys in the store:

```
content_db_oracle_home/jdk/bin/keytool -list -keystore server-keystore.jks
-keypass password -storepass password
```

c. Sign the key before using it:

```
content_db_oracle_home/jdk/bin/keytool -selfcert -validity 5000 -alias
server
-keystore server-keystore.jks -keypass password -storepass password
```

d. Export the server public key from the server keystore to a file:

```
content_db_oracle_home/jdk/bin/keytool -export -alias server -keystore
server-keystore.jks -file cdbServer.pubkey -keypass password -storepass
password
```

e. Store both the keystore password and the private server key password in a secure location so Oracle Content Database can access the keystore and the private key.

```
content_db_oracle_home/content/bin/changepassword -k
```

When prompted for the old password, press [Enter] if it is the first time to set the password; otherwise, enter the previous password. Then, enter and confirm the keystore password (`-storepass password`) that you provided in step 1.b.

See `content_db_oracle_home/content/log/changepassword.log`.

2. Configure a client keystore at the Oracle SES installation.

a. Create the SES client keystore with the following keytool command, substituting a secure password for *password*:

```
ses_oracle_home/jdk/bin/keytool -genkey -keyalg RSA -validity 5000
-alias client -keystore sesClientKeystore.jks -dname "cn=client"
-keypass password -storepass password
```

To list the keys in store:

```
ses_oracle_home/jdk/bin/keytool -list -keystore sesClientKeystore.jks
-keypass password -storepass password
```

b. Sign the key before using the key:

```
ses_oracle_home/jdk/bin/keytool -selfcert -validity 5000 -alias client
-keystore sesClientKeystore.jks -keypass password -storepass password
```

Restart the WebCenter middle tier from the Oracle Enterprise Manager console.

c. Export the server public key from the server keystore to a file:

```
ses_oracle_home/jdk/bin/keytool -export -alias client -keystore
sesClientKeystore.jks -file sesClient.pubkey -keypass password
```

```
          -storepass password
```

3. Import Oracle SES client public keys into the Oracle Content Database server keystore (sesClient.pubkey must be copied to Oracle Content Database):

```
cd content_db_oracle_home/content/settings
```

```
content_db_oracle_home/jdk/bin/keytool -import -alias client -file
sesClient.pubkey -keystore server-keystore.jks -keypass password
-storepass password
```

4. Import Oracle Content Database server public keys into the Oracle SES keystore. (cdbServer.pubkey must be copied to Oracle SES):

```
ses_oracle_home/jdk/bin/keytool -import -alias server -file
cdbServer.pubkey -keystore sesClientKeystore.jks -keypass password
-storepass password
```

> **Note:** Check the server logs at *content_db_oracle_home*/content/logs for keystore issues with the crawler plug-in.

Oracle Content Database Source Attributes

Oracle SES crawls the following attributes for Oracle Content Database Sources:

- AUTHOR

- CREATE_DATE

- DESCRIPTION

- FILE_NAME

- LASTMODIFIEDDATE

- LAST_MODIFIED_BY

- TITLE

- MIMETYPE

- ACL_CHECKSUM: The check sum calculated over the ACL submitted for the document.

- DOCUMENT_LANGUAGE: Oracle SES language code taken from Oracle Content Database language string. For example, if Oracle Content Database uses "American", then Oracle SES submits it as "en-us".

- DOCUMENT_CHARACTER_SET: The character set for the Oracle Content Database document.

Oracle SES also can search categories or customized attributes created by the user in Oracle Content Database.

You can apply categories to files and links, and divide categories into subcategories having one or more attributes. When a document in Oracle Content Database is attached to a category, you can search on the attribute of category. (The attributes appear in the list of search attributes.)

For example, suppose you create a category named testCategory with testAttr1 and testAttr2. Document X is created and assigned to testCategory. You must assign the value to the testCategory attributes. After crawling, testAttr1 and testAttr2 appears in the search attribute list.

Customized attribute values can be the following types: String, Integer, Long, Double, Boolean, Date, User, Enumerated String, Enumerated Integer, and Enumerated Long:

- Index Long, Double, Integer, Enumerated Integer, and Enumerated Long type customized attributes are type Number attributes in Oracle SES. The display name has an _N suffix.

- Index Date customized attributes are type Date attributes in Oracle SES. The display name has a _D suffix).

- Index String, Enumerated String, and User customized attributes are type String attributes in Oracle SES.

Limitations on Custom Attributes for Oracle Content Database

- The Oracle Content Database SDK has more features than the Oracle Content Database Web GUI. The Web GUI does not support String arrays, but the SDK does. If you use the SDK to build customized administration and user GUIs that support the String array type, then a customized attribute can have multiple values.

- If a document in Oracle Content Database is attached to a category and the attributes in that category are left blank, then the attribute is not available in the attribute list for an Advanced Search. The crawler skips attributes with null values. However, if another document has the same attribute with a real value, then the attribute is indexed.

Setting Up Oracle Content Server Sources

The Oracle Content Server connector enables Oracle SES to search Oracle Content Server (formerly Stellent Server), which is the foundation of the Oracle Universal Content Management solution. Users throughout the organization can contribute content from native desktop applications, manage content through rich library services, publish content to Web sites or business applications, and access the content with a browser.

The Content Server connector supports Oracle Content Server 7.5.2 or 10gR3 with XMLCrawlerExport (the Oracle Content Server RSS component).

Oracle Content Server includes an RSS feed generator component (XMLCrawlerExport) on top of the content server. This component generates RSS feeds as XML files from its internal indexer, based on indexer activity. It has access to the original content (for example, a Microsoft Word document), the Web viewable rendition, and all the metadata associated with each document. The component also has a template that contains a Idoc script that applies the metadata values from the indexer to generate the XML document. (Idoc is an Oracle Content Server proprietary scripting language.) Oracle Content Server generates feeds for all documents for the initial crawl, and feeds for updated and deleted documents for the incremental crawl. Each document can be an item in the feed, with the operation on the item (such as insert, delete, update), its metadata (such as author, summary), URL links, and so on.

The Oracle Content Server connector reads the feeds provided by Oracle Content Server according to a crawling schedule. Oracle SES parses and extracts the metadata information, and fetches the document content, using its generic RSS crawler framework.

Oracle SES supports the control feed method, in which individual feeds can be located anywhere and a control feed file is generated containing the links to other feeds. This control file is input to the connector through the configuration file. Control feed must be used when two computers are on different domains or on different platforms, or if

they use remote access protocol, such as HTTP or FTP, for communication between the two servers.

> **See Also:**
>
> - "Overview of XML Connector Framework" on page 4-11
>
> - Oracle Content Database page at
>   `http://www.oracle.com/technetwork/middleware/content-mana`
>   `gement/overview/index.html`

Oracle Content Server Security Model

The Oracle Content Server security model is based on the concept of permissions, which defines the privileges a user has on a document. The following table shows the set of permissions supported by Oracle Content Server. Each permission is a superset of the previous ones. For example, Write permission includes Read permission. Admin permission is a superset of all the permissions.

Table 7–9   Oracle Content Server Permissions

Permission	Description
Read	View documents
Write	View, Check In, Check Out, and Get Copy of documents
Delete	View, Check In, Check Out, Get Copy, and Delete documents
Admin	View, Check In, Check Out, Get Copy, and Delete documents
	An Administration user with Workflow rights can start or edit a workflow for the document. An Administration user can also check in documents with another user specified as the Author.

Oracle Content Server provides multiple security models, including an out-of-the-box security system and integration with centralized security models such as LDAP and Active Directory.

Oracle Universal Content Management security can work in these modes:

- Universal Content Management native identity plugin where Universal Content Management is not connected to a directory

- Oracle Internet Directory

- Active Directory only where Universal Content Management is connected to Active Directory using LDAP. A connection to Active Directory using Microsoft Security is not supported.

The Oracle SES Oracle Content Server connector supports the two most popular security models among current Oracle Content Server customers: Roles and Groups, and Accounts.

Roles and Groups

A security group is a set of files grouped under a unique name. Every file in the library belongs to a security group. Access to security groups is controlled by the permissions, which are assigned to roles, which are assigned to users. For example, the EngAdmin role has Read, Write, Delete, and Admin permission to all content in the EngDocs security group. User Joe is assigned to role EngAdmin; therefore, Joe has all permissions to the documents in EngDocs group.

Accounts

Accounts provide greater flexibility and granularity than groups. An account is a group of content. It introduces another metadata field that is filled out upon content check-in. When accounts are enabled, content items also can be assigned to an account in addition to the security group. A user must have access to the account to read, write, delete or administer content in that account. When accounts are used, the account becomes the primary permission to satisfy before security group permissions are applied.

A user's access to a document is like the intersection between their account permissions and security group permissions. For example, a user is assigned the EngAdmin role, which has all permissions to the documents in EngDocs security group. At the same time, the user is also assigned Read and Write permission to the EngProjA account. Therefore, the user has only Read and Write permission to a content item that is in the EngDocs security group and the EngProjA account.

Accounts can also be set up in a hierarchical structure. A user has permission to the entire subtree starting from the account node. For instance, a user assigned to the Eng account has access to Eng/AbcProj and Eng/XyzProj, or any accounts beginning with Eng. In other words, users that have permission to a particular account prefix also have access to all accounts with that prefix.

> **Note:** Oracle Content Server uses a prefix test for account filtering, so a slash (/) has no special meaning. A user granted permission to account A has access to any documents in account A*, such as A, AB, or A/B. The hierarchical structure takes advantage of the prefix semantics, but it is enforced with the account model. Hence, there is no special character as the level divider when testing for account permissions.

> **See Also:** Oracle Universal Content Management documentation at
>
> http://www.oracle.com/technetwork/middleware/content-managem ent/index-094708.html

Setting Up Identity Management for Oracle Content Server

To activate the Oracle Content Server identity plug-in:

1. On the Global Settings page, select **Identity Management Setup** under the System heading.

    The Global Settings - Identity Management Setup page is displayed.

2. Select **Oracle Content Server** and click **Activate**.

3. Enter values for the parameters described in Table 7–10, then click **Finish**.

Table 7–10   Oracle Content Server Connector Setup Parameters

Parameter	Value
HTTP endpoint for authentication	HTTP endpoint for Oracle Content Server authentication. For example, http://my.host.com:*port*/idc/idcplg
Admin User	Administrative user who accesses the Oracle Content Server Identity Service API
Password	Administrative user password

Creating an Oracle Content Server Source

To create an Oracle Content Server source using the Oracle SES Administration GUI:

1. On the home page, click the **Sources** secondary tab to display the Sources page.

2. Select **Oracle Content Server** from the **Source Type** list, then click **Create** to display Step 1 Parameters.

3. Enter values for the parameters described in Table 7–11.

4. Click **Next** to display Step 2 Authorization, then set values for the parameters described in Table 7–11.

5. Scroll down to Security Attributes to verify that `ACCOUNT` and `DOCSECURITYGROUP` are listed. If they are not, then the source was not created correctly. Verify that the Configuration URL in Step 1 is correct.

6. Click **Create** to create the Oracle Content Server source.

   After processing each data feed, a status feed is uploaded to the location specified in the configuration file. This status feed is named one of the following:

   - *data_feed_file_name*.suc indicates the data feed was processed successfully.

   - *data_feed_file_name*.err indicates that an error was encountered while processing the feed. The errors are listed in this status feed.

   Tip: To index multibyte character sets, set the default character set of the crawler to UTF-8 regardless of the character set of Oracle Content Server. See "Modifying the Crawler Parameters" on page 4-2.

Table 7–11   Oracle Content Server Source Parameters (Step 1)

Parameter	Value
Configuration URL	URL of the XML configuration file providing details of the source, such as the data feed type, location, security attributes, and so on. Obtain the location of the file from the Oracle Content Server administrator.
	Use the following format to enter the configuration URL:
	`http://`*host_name*`/`*instance_name*`/idcplg?IdcService=SES_` `CRAWLER_DOWNLOAD_CONFIG&source=`*source_name*
Authentication Type	Java authentication type. Set this parameter when the data feeds are accessed over HTTP.
	Enter one of the following values:
	- `NATIVE`: Proprietary XML over HTTP authentication
	- `ORASSO`: Oracle Single Sign-on.
User ID	User ID to access the data feeds. The access details of the data feed are specified in the configuration file. Obtain a user ID from the Oracle Content Server administrator.
Password	Password for **User ID**. Obtain the password from the Oracle Content Server administrator.
Realm	Realm of the Oracle Content Server instance. The value for the Realm field is required only when the authentication type is set to either `Basic` or `Digest`. As Oracle Content Server sources use `Native` authentication, the value for the Realm field must be left blank.

Table 7–11 (Cont.) Oracle Content Server Source Parameters (Step 1)

Parameter	Value
Oracle SSO Login URL	Set this parameter value when the authentication type is ORASSO. Oracle SES redirects the crawler to this SSO login URL for authentication before crawling the Oracle Content Server source.
	The format for this URL for Oracle 10g SSO secured Oracle Content Server source is:
	■ For basic authentication type:
	`https://server:port/pls/orasso/orasso.wwsso_app_admin.ls_login`
	■ For form based authentication type:
	`https://server:port/mysso/signon.jsp`
	The format for this URL for Oracle 11g OAM secured Oracle Content Server source is:
	`http://server:port/oam/server/obrareq.cgi?encquery`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.
Oracle SSO Action URL	Set this parameter value when the authentication type is ORASSO. This is the URL that is displayed after successfully authenticating the SSO user before crawling the Oracle Content Server source.
	The format for this URL for Oracle 10g SSO secured Oracle Content Server source is:
	`https://server:port/sso/auth`
	The format for this URL for Oracle 11g OAM secured Oracle Content Server source is:
	`http://server:port/oam/server/auth_cred_submit`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.
Scratch Directory	Directory where Oracle SES can write temporary status logs. The directory must be on the same system where Oracle SES is installed. Optional.
Maximum number of connection attempts	Maximum number of attempts to connect to the target server for access to the data feed.
Delete Linked Document	Set to `true` if the documents crawled from the links in the feeds must be deleted, else set to `false`.
Number of data feeds to be pre-fetched	Number of data feeds to pre-fetch. Its value can be any number greater than or equal to 0. If its value is set to 0, then no data feed is pre-fetched. The value specified must be such that there is sufficient memory to cache the pre-fetched feeds. The default value is 0.
Enable Resume Crawl	Set it to `true` to enable resuming crawl when the crawl fails or stops. The default value is `true`.
Stop Crawl On Content Fetch Error	Set it to `true` to stop crawl on content fetch error, else set it to `false`. The default value is `false`.

Table 7–12 Oracle Content Server Connector Authorization Parameters (Step 2)

Parameter	Value
HTTP Endpoint for Authorization	HTTP endpoint for Oracle Content Server authorization, such as `http://example.com:7777/idc/idcplg`.

Table 7–12   (Cont.) Oracle Content Server Connector Authorization Parameters (Step 2)

Parameter	Value
Display URL Prefix	HTTP host information to prefix the partial URL specified in the access URL of the documents in RSS feeds to form the complete URL. This complete URL is displayed as the URL when a user clicks the document link in the Oracle SES search results page. For example, you might display `http://example.com:7777/idc` (*not* `http://example.com/`, as shown on the user interface page).
Administrator User	Administrative user to access the Authorization Service API of Oracle Content Server.
Administrator Password	Administrative user password.
Display Crawled Version	Controls access to the crawled documents: ■ `true`: Search results point to the crawled version of the document. ■ `false`: Search results point to the content information page.
Authorization User ID Format	Format of the user ID used by the Oracle Content Server authorization API, such as `username`, `email`, `nickname`, `user_name`.
Use Cached User and Role Information to Authorize Results	Controls user authorization: ■ `true`: Uses the cached user query filter. This setting removes the query time dependency on Oracle Content Server. ■ `false`: Queries Oracle Content Server for authorization.
User Role Data Source to Cache the Filter	The name of the Oracle Content Server Users source that has crawled the user's SecurityGroup and Account information.

8

Configuring Access to Collaboration Sources

This chapter contains the following topics:

- Setting Up Lotus Notes Sources
- Setting Up Microsoft Exchange Sources
- Setting Up NTFS Sources for Windows
- Setting Up NTFS Sources for Linux and UNIX
- Setting Up Oracle Calendar Sources
- Setting Up Oracle Collaboration Suite E-Mail Sources

Setting Up Lotus Notes Sources

Lotus Notes data is stored in notes-databases, which can be further contained inside directories on a server. A Lotus Domino Server instance can have one or more databases that can be crawled using the Lotus Notes source. The Lotus Notes source navigates through the databases to crawl the documents (for example, e-mail, calendar, address book, and "to do") in the specified databases. It stores the metadata, and accesses information in Oracle SES to provide search according to the end users' credentials.

The Lotus Notes connector now lets you enable or disable multiple attachment support with the `Attachment as Search Item` attribute. When this is disabled, the additional attributes `Parent URL` and `Parent Title` are added for all attachment documents, to link it with the parent document.

The Lotus Notes source supports incremental crawling; that is, it crawls and indexes only those documents that have changed after recent most crawling was scheduled. A document is re-crawled if either the content, metadata, display URL or the direct security access information of the document has changed. Documents deleted from a database are removed from the index during incremental crawling.

To enable Oracle SES to launch Notes thick client, set the **Notes Thick Client** parameter to `true`.

Important Notes for Lotus Notes Sources

The user-account used to crawl Lotus Notes databases should preferably be an Administrator account, such that it has access on all databases and can retrieve and crawl all documents in the specified databases.

Required Software

- Lotus Domino Server R5.0.9/R6.5.4/R7.0

- Notes Clients R5.0.9/R6.5.4/R7.0

Required Tasks

The following tasks must be performed before installing the Lotus Notes source:

1. HTTP and DIIOP tasks must be running on Domino Server.

2. If the Active Directory identity plug-in is used, then the users and user-groups in the Domino Directory must be synchronized with Active Directory. While using the Active Directory identity plug-in, the short-name in the Lotus Notes person document is used for validating the user in Active Directory, so it should be a resolvable logon name in Active Directory.

3. Configure the server document:

    a. Open the server document on the Lotus Notes server that must be crawled.

    b. On the Configuration page, expand the **Server** section.

    c. On the Security page, in the **Programmability Restrictions** area, specify the appropriate security restrictions for your environment in the following fields:

    Run restricted Lotus Script/Java agents

    Run restricted Java/Javascript/COM

    Run unrestricted Java/Javascript/COM

    For example, you might specify an asterisk (*) to allow unrestricted access by Lotus Script/Java agents, and specify user names that are registered in the Domino Directory for the Java/Javascript/COM restrictions.

    Note: The crawler that you configure to crawl this server with the DIIOP protocol must be able to use the user names that you specify in these fields.

    ---

    d. Open the Internet Protocol page, then open the HTTP page, and set the **Allow HTTP Clients to Browse Database** option to **Yes**.

    e. Configure the user document:

    Open the user document on the Lotus Notes server. This document is stored in the Domino directory.

    On the Basics page, for **Internet password**, specify a password.

    f. Restart the DIIOP task on the server.

4. Copy the Lotus Notes JAR files `Note.jar` and `NCSO.jar` to the directory *ses_home*`/search/lib/plugins/ln/` before activating the Lotus Notes identity plug-in.

Known Issues

- A Lotus Notes source does not index encrypted fields, and the content of attachments with encrypted documents, for searching. With encrypted documents, the URL of the search result launches the Notes document instead of the attachment file, which is the case when non-encrypted documents are crawled.

■ Deleted Notes documents and attachments in Notes documents are still searchable after an incremental crawl that was set by specifying 'Recrawl using last modified date' as true. To remove URLs from deleted documents or attachments from the Oracle SES index, either perform a force re-crawl (that is, change the re-crawl policy to **Process All Documents** on the Home - Schedules - Edit Schedule page) or mark the 'Recrawl using last modified date' source parameter as false.

Setting Up Identity Management for Lotus Notes

Activate an identity plug-in on the Global Settings - Identity Management Setup page.

The users/groups on Active Directory can be synchronized with Lotus Domino Directory such that all users/groups in Active Directory get registered in Domino as well. Thus, any ACL entry in a notes database or notes document can be validated in Active Directory also, and vice versa.

> **See Also:** "Activating the Active Directory Identity Plug-in" on page 11-10

Oracle SES also provides a Lotus Notes identity plug-in so the Lotus Domino Directory can be used to authenticate and validate the notes native users and groups in Oracle SES.

Activate the Lotus Notes identity plug-in with the following parameters:

■ **Server name**: The Domino server fully qualified host name/IP address. If the HTTP port on the Domino server is not 80, then the host name should be *server_name:HTTP_port_number*.

■ **User name**: User name of a valid Lotus Domino Server user. Required.

■ **Password**: Internet password of the Lotus Notes user. Required.

Creating a Lotus Notes Source

Create a Lotus Notes source on the Home - Sources page. Select **Lotus Notes** from the Source Type list, and click **Create**. Enter values for the following parameters:

■ **Server Name**: The Domino server fully qualified host name/IP address. For example, if the Lotus Notes database name is `ses.nsf`, then enter `ses.nsf` for this parameter. If the HTTP port on the Domino server is not 80, then the host name should be *server_name:HTTP_port_number*. Required.

■ **Attribute list**: The comma-delimited list of Lotus Notes attributes along with their data types to search. The format is *AttributeName:AttributeType, AttributeName:AttributeType*. The valid values are String, Number, and Date. For example: `Subject:String`

Table 8–1   Lotus Notes Data Type Mapping

Sr. No	Lotus Notes Data Type	Oracle SES Data Type
1	Boolean	String
2	Integer	Number (Big Decimal)
3	String	String
4	Date	Date

While crawling a database, an attribute is indexed only if both name and type match the configured name and type; otherwise, it is ignored. This is an optional parameter.

The default searchable attributes for Lotus Domino Server are LASTMODIFIEDDATE, Title, and Author. Multiple attributes with same name are not allowed.

- **User name**: The user name of a valid Lotus Domino Server user. The user should be an Administrator user or a user who has access to all folders and documents of the databases configured in the **Container name** parameter. The user should be able to retrieve content, metadata, and ACL from documents of all databases configured in **Container name** parameter. Required.

- **Password**: Internet password of the Lotus Notes user. Required.

- **Container Name**: Names of the containers to be crawled. Multiple container names must be comma-delimited. The container name can include folders, databases, views and folders within databases. For example, `database-abc.nsf`, `folders-folder1`, `views-abc.nsf:By Author`, and `db-abc.nsf:folder\subfolder`. Note that Lotus Notes database file name must be specified with the extension.

- **Crawl Public Documents**: Indicate whether the public documents on notes databases must be crawled such that they are available to anonymous users in Oracle SES, either true or false. Required.

- **Authentication Attribute**: The attribute used to validate the ACL. With the Active Directory identity plug-in, set the value to `USER_NAME`. With the Lotus Notes identity plug-in, set the value to `NATIVE`. Required.

- **Mail Template Name**: This parameter is specific to the mail-databases and the mail template's name should be specified here if any/all of the databases being crawled are mail databases. This is a mandatory parameter if either the **Past Days** or **Future Days** parameter is specified.

- **Past Days**: If the user is crawling calendar entries, then this parameter specifies the number of days in the past for which the calendar entries are picked. The date of reference here is the start date of the event. This accounts for the number of days in the past, and it does not filter the search by time.

- **Future Days**: If the user is crawling calendar entries, then this parameter specifies the number of days in the future for which the calendar entries are picked. The date of reference here is the end date of the event. This accounts for the number of days in the future, and it does not filter the search by time.

- **Notes Title Field**: Because in Lotus Notes custom applications it is not mandatory to maintain a Title field, this parameter has been provided to specify those text fields that should be parsed to retrieve the title field. For example, you could enter `Subject`. With multiple field names, the first field available on the document is selected for the title. Required.

- **Notes Thick Client**: Enter `true` to use Lotus Notes (thick client). Enter `false` to use Lotus Notes Web access.

- **Recrawl using last modified date**: Enter `true` to enqueue only modified documents. Required.

- **Attachment As Search Item**: Enter `true` to have each document in the attachment be submitted individually as an independent document with the same set of attributes and ACLS as that of the parent document. Enter `false` to have attachments be added to the parent document and submitted as a single unit.

Displaying the Parent URL in the Search Results

Take the following steps to display the Parent URL attribute in the search results for Lotus Notes connector:

1. On the Global Settings page, select **Use Advanced Configuration**. The Global Settings - Configure Search Result List is displayed.

2. Under Attribute Selection, move **ParentURL** to the **Included** list.

3. Under Style Sheets in the Enter an XSLT box, scroll to `<!-- Links link -->...</td>` and enter the following XSL code:

```
<tr>
   <td>
      <xsl:if test="parenturl [.!= '']">
         <xsl:text>ParentURL: </xsl:text>
         <a class="browseLink" href="{parenturl}">
            <xsl:value-of select="parenturl" />
         </a>
      </xsl:if>
   </td>
</tr>
```

4. Click **Apply**.

Setting Up Microsoft Exchange Sources

Oracle SES can crawl through and provide secure search for e-mail and calendar items, related metadata, attributes, ACLs, and attachments in Microsoft Exchange. It also provides attribute search and browse functionality, which allows search to be done against a specific subfolder in the hierarchy.

Oracle SES supports incremental crawling; that is, it crawls and indexes only those documents that have changed since the last crawl was scheduled. A document is re-crawled if either the content or metadata or the direct security access (permissions) information of the document has changed. A document is also re-crawled if it is moved within Microsoft Exchange. Documents deleted from Exchange are removed from the index during incremental crawls.

A Microsoft Exchange source covers the following objects in Exchange:

- E-mail
- E-mail attachments
- Calendar events

Important Notes for Microsoft Exchange Sources

On the Exchange server, the super user must grant himself the `Send as` and `Receive as` privileges. You can enable privileges globally for all users in the system. No user-specific privilege grants are required.

See Also:

- *Microsoft Exchange 2003 Technical Reference Guide* and information about permissions in Microsoft Exchange: http://technet.microsoft.com/en-us/library/aa996429%28v=exchg.65%29

- *Oracle Secure Enterprise Search Release Notes* on OTN for supported platforms

Required Software

- Microsoft Internet Information Server (IIS)

Note: The file ADODB.dll is usually included in the Microsoft .NET Framework SDK. However, if this file is not on your computer, then you must download the ADODB.dll appropriate for your system from Microsoft and install it using the following command:

```
gacutil /i adodb.dll
```

You can download the Microsoft .NET Framework from this site:

http://www.microsoft.com/en-us/download/default.aspx

Required Tasks

- Proper permissions on the Exchange server must be granted to the Exchange administrator. The Exchange server is crawled with the permission of a super user with the Send as and Receive as privileges. The easiest way to configure this is to use an administrator as super user or create a super user with the administrator privilege and the Send as and Receive as privileges targeting Exchange inbox store and public folders.

- To enable the Outlook Web Access logon page, you must enable forms-based authentication on the server. To enable forms-based authentication:

  1. On the Exchange server, log on with the Exchange administrator account, and then start Exchange System Manager.

  2. In the console tree, expand **Servers**.

  3. Expand the server for which you want to enable forms-based authentication, and then expand **Protocols**.

  4. Expand HTTP, right-click **Exchange Virtual Server**, and then click **Properties**.

  5. In the **Exchange Virtual Server Properties** dialog box, on the **Settings** tab, in the **Outlook Web Access** pane, select the **Enable Forms Based Authentication** option.

  6. Click **Apply**, and then click **OK**.

  7. Restart the IIS server.

  If you are using forms-based authentication with SSL off-loading, you must configure your Exchange Server front-end servers to handle this scenario.

  See Also: *How to Enable Forms-Based Authentication* at

  http://technet.microsoft.com/en-us/library/bb123832.aspx

Known Issues

E-mails with multibyte characters sent from a browser with a different language set than the characters in the mail are not indexed correctly in Oracle SES. The multibyte characters are converted to question marks (?).

This is a known e-mail content issue with Microsoft Exchange. To send future e-mails so that the Microsoft Exchange connector can crawl them properly, either of these workarounds can be applied:

- Change the browser language to the characters in the e-mail. For example, set it to "Japanese" to input Japanese characters.

- Change the value of the following registry key:

```
HKEY_LOCAL_
MACHINE\SYSTEM\CurrentControlSet\Services\MSExchangeWEB\OWA\UseRegionalCharset
     (Original) '1'
     (New)      Any number (except 1). For example, '0'
```

> **See Also:**
>
> - *How to Modify the Default Browser Language Settings for Outlook Web Access* at
>
>   http://technet.microsoft.com/en-us/library/aa996640.aspx
>
> - *Outlook Mobile Access and Exchange 2003* at
>
>   http://technet.microsoft.com/en-us/library/dafc9621-7b71-42fa-b1cb-3ea63e15ad04.aspx

Setting Up Identity Management for Microsoft Exchange

The Microsoft Exchange connector uses WebDAV for best performance. Oracle recommends that Active Directory be used as identity management system for the Oracle SES instance. The Active Directory instance must be the same one that Microsoft Exchange is using to authenticate users on the file system.

For the Oracle SES instance to read the files during crawling, add permission to each folder and file to make them accessible by the operating system user that runs the Oracle SES instance. Adding permissions to a folder automatically adds the same permissions to all the files and subfolders in the folder.

> **See Also:** "Activating the Active Directory Identity Plug-in" on page 11-10

Creating a Microsoft Exchange Source

Create a Microsoft Exchange source on the Home - Sources page. Select **Microsoft Exchange** from the Source Type list, and click **Create**.

Enter values for the following parameters:

- **User Name**: User name to authenticate between Oracle SES and Exchange

- **Password**: password to authenticate between Oracle SES and Exchange

- **Server**: Microsoft Exchange server IP

- **Domain**: Microsoft Exchange server domain

- **LDAP Port**: Microsoft Exchange LDAP port

- **Simple Include**: To limit crawling, specify up to 50 colon-delimited path inclusion boundary rules using simplified regular expressions. Specify an inclusion rule that a URL contain, start with, or end with a term. Only `*`, `^`, and `$` operators are permitted. An asterisk (`*`) is a wildcard. A caret (`^`) denotes the beginning of a URL, and a dollar sign (`$`) denotes the end. For example:
  ```
  ^https://*.oracle.com/.jpg$
  ```

- **Simple Exclude**: To limit crawling, specify up to 50 colon-delimited path exclusion boundary rules using simplified regular expressions. Only `*`, `^`, and `$` operators are permitted.

- **Regular Expression Include**: To limit crawling, specify up to 50 colon-delimited path inclusion boundary rules using restricted (full java.util.regexp) regular expression rules. For example:
  ```
  ^https://.*\.oracle(?:corp){0,1}\.com
  ```

- **Regular Expression Exclude**: To limit crawling, specify up to 50 colon-delimited path exclusion boundary rules using restricted (full java.util.regexp) regular expression rules.

Microsoft Exchange Source Attributes

- ReceivedTime
- From
- To
- CC
- Subject
- Lastmodifieddate

Setting Up NTFS Sources for Windows

This section contains information for NTFS sources on Windows. For NTFS on UNIX, see "Setting Up NTFS Sources for Linux and UNIX" on page 8-10.

The NTFS connector enables Oracle SES to search file repositories in Microsoft NTFS. An Oracle SES NTFS source collects the content, metadata attributes and ACLs of files in NTFS. An NTFS source supports incremental crawl. After the initial crawl is performed, subsequent crawls only collect those documents that have changed since the last crawl. A document is re-crawled if the content, metadata, or the ACL information of the document has changed. A file is also re-crawled if it is moved between folders. Files deleted from NTFS are removed from the index during incremental crawls.

Important Notes for NTFS Sources

- The operating system user running the Oracle SES instance must have read permission on the NTFS file share being crawled. For example, if the remote file share `\\computer1\share1\directory1\` is crawled by the NTFS source, then the Oracle SES instance must be run as a domain user who has access to the file share.

- If you get the ACL in the form `<encrypted acl>@domain` for a folder on a remote computer, it probably means that the computer running the Oracle SES instance and the remote computer are on different domains and your computer cannot interpret the ACLs appropriately.

- Currently, the Oracle SES crawler considers the shared folder an empty document, but it is not indexed; therefore, the total number of unique documents indexed is less than the total number of documents fetched.

- An ACL error may appear when crawling an NTFS source as a built-in user or group, such as an Administrator user. As a workaround, set explicit access to the administrator user by right clicking the NTFS share and adding the administrator user with all the permissions.

- *Everyone* is a special group that represents all current network users, including guests and users from other domains. When a user logs on to the network, the user is automatically added to the *Everyone* group. The NTFS connector supports the *Everyone* group. All documents for which the *Everyone* group has permission is crawled and accessed like public documents. There is no need to log in to the query application to access these public documents. However, if there is a *deny* to a user along with permissions to *Everyone* group to access the document, then all users except for the one for who *deny* has been granted can see the document, and these users must log in to the query application to see the document.

- When using Internet Explorer with files on a different domain, you must explicitly log in to Internet Explorer to open result links to those files.

- When you use the NTFS connector and search file types of `.txt`, `.zip`, or `.rtf`, only the Title and Author attributes are fetched and indexed. For these attributes, the crawler fetches the properties stored in the authoring program (typically accessed by selecting **Properties** from the File menu) and not the NTFS properties (accessed in Windows Explorer by right-clicking the file name and choosing **Properties**).

Required Software

Microsoft .NET Framework 2.0

Required Tasks

If not previously installed, then download and install the Microsoft .NET 2.0 Framework from this site:

http://www.microsoft.com/en-us/download/default.aspx

Setting Up Identity Management for NTFS Sources

When an NTFS source is used, Oracle recommends that Active Directory be used as identity management system for the Oracle SES instance. The Active Directory instance must be the same one that NTFS is using to authenticate users on the file system.

For the Oracle SES instance to read the files during crawling, add the permission to each folder and file to make it accessible by the operating system user that runs the Oracle SES instance. Adding permissions to a folder automatically adds the same permissions to all the files and sub-folders in the folder.

NTFS sources rely on Active Directory for security permissions. Because permissions at the server local group level are not defined in Active Directory, these permissions are not supported when crawling NTFS sources. Permissions for server local groups (not domain local groups) are ignored during crawling. Permissions for domain groups and users inherited from server local groups also are ignored.

> **See Also:** "Activating the Active Directory Identity Plug-in" on page 11-10

Creating an NTFS Source

Create an NTFS source on the Home - Sources page. Select NTFS from the Source Type list, and click **Create**. Enter values for the following parameters:

- **UNC Path**: UNC Paths, for example, `\\MyServer\Mysharedfolder`

- **Domain Name**: Domain name of the URL (**UNC Path**)

- **Simple Include**: To limit crawling, specify up to 50 colon-separated path boundary rules using simplified regular expressions. Only `*`, `^`, and `$` operators are permitted. For example: `^https://*.oracle.com/.jpg$`

- **Simple Exclude**: To limit crawling, specify up to 50 colon-separated path boundary rules using simplified regular expressions. Only `*`, `^`, and `$` operators are permitted.

- **Regular Expression Include**: To limit crawling, specify up to 50 colon-separated path boundary rules using restricted (full java.util.regexp) regular expression rules. For example: `^https://.*\.oracle(?:corp){0,1}\.com`

- **Regular Expression Exclude**: To limit crawling, specify up to 50 colon-separated path boundary rules using restricted (full java.util.regexp) regular expression rules.

- **Use Local Display URL**: Enter `true` to use the local display URL or `false` to use display the content in a web browser.

- **Authentication Attribute**: Authentication attribute used by the LDAP to validate the user. Use `USER_NAME` for Active Directory and `nickname` for Oracle Internet Directory.

After crawling an NTFS source, you may get a "No User Found Matching the Criteria" error message on the Home - Schedules - Data Synchronization page. This error is signalled by the identity plug-in. The NTFS connector tries to validate the principal as user first. If that fails, then it tries to validate the principal as group. This error occurs if there are groups as ACL for a document, because the connector does not know if the given principal is a user or a group.

NTFS Source Attributes

- ACLS_
- FILEDATE
- Host
- Language
- LastModifiedDate
- Mimetype
- Title

Setting Up NTFS Sources for Linux and UNIX

This section contains information for NTFS sources on Linux and UNIX platforms, which have additional setup steps not required on Windows. For NTFS sources on Windows, see "Setting Up NTFS Sources for Windows" on page 8-8.

An NTFS source collects the content, metadata attributes, and ACLs of files in NTFS. An NTFS source supports incremental crawl. After the initial crawl is performed,

subsequent crawls only collect those documents that have changed since the last crawl. A document is re-crawled if the content, metadata or the ACL information of the document has changed. A file is also re-crawled if it is moved between folders. Files deleted from NTFS are removed from the index during incremental crawls.

Important Notes for NTFS Sources

- On the Windows server, the super user must have permission to read the NTFS file share.

- The super user must be the impersonate user in the IIS Server.

- The default behavior for NTFS for Linux and UNIX platforms is to use local file display URL, so the client computer must have access to the file share.

- An ACL error may appear when crawling an NTFS source as a built-in user or group, such as an Administrator user. As a workaround, set explicit access to the administrator user by right clicking the NTFS share and adding the administrator user with all the permissions.

- *Everyone* is a special group that represents all current network users, including guests and users from other domains. When a user logs on to the network, the user is automatically added to the Everyone group. The NTFS connector supports the Everyone group. All documents for which the Everyone group has permission is crawled and accessed like public documents. There is no need to log in to the query application to access these public documents. However, if a user is denied access to a document while the Everyone group has access, then all users except for the denied user can see the document, and these users must log in to the query application to see the document.

- When using Internet Explorer with files on a different domain, you must explicitly log in to Internet Explorer to open result links to those files.

Required Software

- Microsoft Internet Information Server (IIS)

- Microsoft .NET 2.0 Framework

Setting Up Identity Management with NTFS Sources

When an NTFS source is used, Oracle recommends that Active Directory be used as identity management system for the Oracle SES instance. The Active Directory instance must be the same one that NTFS is using to authenticate users on the file system.

For the Oracle SES instance to read the files during crawling, add permission to each folder and file to make them accessible by the operating system user that runs the Oracle SES instance. Adding permissions to a folder automatically adds the same permissions to all the files and sub-folders in the folder.

NTFS sources rely on Active Directory for security permissions. Because permissions at the server local group level are not defined in Active Directory, these permissions are not supported when crawling NTFS sources. Permissions for server local groups (not domain local groups) are ignored during crawling. Permissions for domain groups and users inherited from server local groups also are ignored.

> **See Also:** "Activating the Active Directory Identity Plug-in" on page 11-10

Creating an NTFS Source Type on Linux and Unix Platforms

To create an NTFS source type on Linux or UNIX platform:

1. On the Global Settings page, click the **Source Types** link.

2. On the Source Types page, click **Create**.

3. Provide the following parameter values on the Complete the New Source Type : Step 1 page:

Paramater	Value
Name	`Microsoft NTFS (Linux and Unix)`
Plug-in Manager Java Class Name	`oracle.search.plugin.ntfs.linux.NTFSLinuxCrawlerPluginManager`
Plug-in Jar File Path	`ntfs/ntfs.jar`

4. Click **Next**.

5. On the Complete the New Source Type : Step 2 page, click **Finish**.

Now, create the NTFS source on Linux or Unix platform as described in the following section.

Creating an NTFS Source for Linux and Unix Platforms

To create an NTFS source on Linux or UNIX platform:

1. On the home page, select the **Sources** secondary tab.

2. On the Sources page, select the **NTFS** source type and click **Create**.

3. Complete the Create User-Defined Source page. Table 8–2 describes the parameters.

4. Click **Create** or **Create & Customize**.

Table 8–2   NTFS Source Parameters for Linux and UNIX Platforms

Parameter	Description
UNC Path	UNC path for the NTFS system to crawl; for example, `\\MYSERVER\mysharedfolder`
WebService Endpoint	Target end point (HTTP or HTTPS); for example `https://mail.example.com/NTFSWebService/NTFSWebService.asmx`
WebService User Name	User name to authenticate the NTFS WebService for the Endpoint.
WebServicePassword	Password for **User Name**.
Simple Include	To limit crawling, specify up to 50 colon-delimited (:) path inclusion boundary rules using simplified regular expressions. Specify an inclusion rule that a URL contain, start with, or end with a term. Only *, ^, and $ operators are permitted. An asterisk (*) is a wildcard. A caret (^) denotes the beginning of a URL, and a dollar sign ($) denotes the end of a URL. For example: `^https://*.oracle.com/.jpg$`
Simple Exclude	To limit crawling, specify up to 50 colon-delimited (:)path exclusion boundary rules using simplified regular expressions. Only *, ^, and $ operators are permitted.

Table 8–2   (Cont.) NTFS Source Parameters for Linux and UNIX Platforms

Parameter	Description
Regular Expression Include	To limit crawling, specify up to 50 colon-delimited (:) path inclusion boundary rules using restricted (full java.util.regexp) regular expression rules. For example: `^https://.*\.oracle(?:corp){0,1}\.com`
Regular Expression Exclude	To limit crawling, specify up to 50 colon-delimited (:) path exclusion boundary rules using restricted (full java.util.regexp) regular expression rules.
ACL Validation Attribute	ACL attribute used to validate the user. Enter `USER_NAME` for Active Directory or `nickname` for Oracle Internet Directory.
Domain Name	Domain name of the URL (UNC Path).
Incremental Crawl With File Change Detector	Enter `true` to use the File Change Detector, or `false` to use scan-based incremental crawl. See "Installing Oracle Search File Change Detector" on page 8-13

After crawling an NTFS source, you may get a "No User Found Matching the Criteria" error message on the Home - Schedules - Data Synchronization page. If this error accompanies a crawl failure, then check that the principal is a valid user or group.

Installing and Configuring Windows Services

NTFS sources on Linux and UNIX platforms requires an NTFS agent to be installed and configured on the Windows domain where the NTFS files are to be crawled. The NTFS agent collects and sends content and metadata to the crawler plug-in on the Oracle SES computer in a crawl session. The communication protocol between Oracle SES and the NTFS agent is HTTP or HTTPS.

The NTFS agent must be installed on a Windows computer where IIS is present, and the computer must be in the same Windows domain where the NTFS file share to be crawled resides.

Typically, a remote file share is crawled with the permission of a domain administrator or a domain user with read privileges on the file share. The easiest way to configure this is to add the domain admin group to the 'administrators' group of the target computer.

The Oracle SES instance must connect to the same Active Directory instance that the Microsoft NTFS domain connects to.

Required Software

Microsoft .NET Framework 2.0

Internet Information Services (IIS) Manager

Required Tasks

Verify that Microsoft .NET 2.0 Framework is installed. If it is not, then download and install it from this site:

http://www.microsoft.com/en-us/download/default.aspx

Installing Oracle Search File Change Detector

By installing and configuring the Oracle Search File Change Detector service, you can realize significantly improved performance in incremental crawls. This service

provides the crawler with a list of documents that are modified or deleted. This method is more efficient than scanning all files for changes.

The older, scan-based incremental crawl is still available and can be used when File Change Detector cannot be deployed on your NTFS system or under the conditions listed in "Configuring the NTFS Connector" on page 8-19.

The following procedure installs File Change Detector in the Microsoft .NET Framework.

To install Oracle Search File Change Detector:

1. Copy `OracleSearchFileChangeDetector.zip` from *ses_home*`/search/lib/plugins/ntfs` to the Windows server where Internet Information Services (IIS) is running.

2. Unzip the contents of `OracleSearchFileChangeDetector.zip` file to a folder. This zip file contains two files:

   - `OracleSearchFileChangeDetector.exe`

   - `OracleSearchFileChangeDetector.exe.config`

3. Open `OracleSearchFileChangeDetector.exe.config` in a text editor and modify the configuration settings as necessary. The settings are described in "Modifying the File Change Detector Configuration File" on page 8-14.

4. Open a command prompt window and navigate to the folder for .NET Framework Version 2.0. For example:

   `C:\WINDOWS\Microsoft.NET\Framework\v2.0.50727`

5. Install the **OracleSearchFileChangeDetector** service by issuing a command like the following, where *path* is the directory containing the configuration file:

   `InstallUtil path\OracleSearchFileChangeDetector.exe`

   For example:

   `installutil d:\OracleSearchFileChangeDetector\`
   `OracleSearchFileChangeDetector.exe`

   The **Set Service Login** dialog box is displayed.

6. Enter the user credentials for the domain user identified in the ASP.NET Configuration Settings dialog box. For **Username**, use the format *domain\username*.

7. Open the Windows Services utility and start the **OracleSearchFileChangeDetector** service.

8. Install the NTFS Web service as described in "Installing the NTFS Web Service" on page 8-16.

Modifying the File Change Detector Configuration File

The `OracleSearchFileChangeDetector.exe.config` file is the XML configuration file for the File Change Detector. When you add new sources, this file is automatically updated with the UNC path of the sources. However, if you make changes to the path of an already existing source, then you must restart File Change Detector for the new path to be watched.

Example 8–1 shows a sample configuration file.

Example 8–1   Oracle Search File Change Detector Configuration File

```
<?xml version="1.0" encoding="utf-8" ?>
<configuration>
   <configSections>
      <section name="StartupFolders"
         type="FileChangeDetector.StartupFoldersConfigSection,
         OracleSearchFileChangeDetector"/>
   </configSections>
   <StartupFolders>
      <DefaultInternalBufferSizeValue>
         <add internalBufferSize="32768" />
      </DefaultInternalBufferSizeValue>
      <Folders>
         <add sourceName="NTFS1" path="10.255.255.255\writeHere"/>
         <add sourceName="NTFS2" path="10.255.255.255\Work"
            internalBufferSize="40960"/>
      </Folders>
      <Results>
         <add directory="C:\NTFS\Data" />
      </Results>
      <SESBufferSizeValue>
         <add sesBufferSize="1" />
      </SESBufferSizeValue>
   </StartupFolders>
</configuration>
```

The XML elements are described in the following topics.

DefaultInternalBufferSizeValue

Oracle Search File Change Detector uses a Windows API to capture file update events. The API uses an internal buffer to cache events. The buffer size is specified in the internalBufferSize parameter of the nested add element:

```
<DefaultInternalBufferSizeValue>
   <add internalBufferSize="n" />
</DefaultInternalBufferSizeValue>
```

The internalBufferSize parameter specifies the default buffer size for all folders that the File Change Detector monitors, as specified in the Folders element.

The internal buffer is allocated from non-paged memory, which cannot be swapped to disk. Therefore, keep the value of internalBufferSize as small as possible. Increase the value for frequent, highly concurrent updates: More than 100 changes per second.

Folders

This element specifies the list of directories to be watched. Create one nested add element for each NTFS source:

```
<Folders>
   <add sourceName="name" path="path"/>
   <add sourceName="name" path="path" internalBufferSize="n"/>
</Folders>
```

The nested add element has these attributes:

- **SourceName**: A unique name within the configuration file to identify the NTFS source. (Required)

- **Path**: The UNC path specified in the NTFS source configuration. (Required)

  To specify multiple UNC paths, use colon as the delimiter. For example:

  ```
  <add sourceName="ntfstest" path="\\server1\share1\Folder:\\server2\share1"/>
  ```

- **InternalBufferSize**: A value that overrides DefaultInternalBufferSizeValue for a source where extensive changes are expected. (Optional)

Results

Specifies the folder where the Oracle Search File Change Detector logs the changes. The value must be the same as the IncrementalCrawlData property in the Web service configuration.

```
<Results>
   <add directory="path" />
</Results>
```

SESBufferSizeValue

This element specifies the number of events cached in an internal buffer by the OracleSearchFileChangeDetector service before writing them to the log file. For example, a value of 1 indicates that every event is written immediately to the log file, while a value of 10 means that 10 events are cached before writing them to the log file.

Increase the value of the sesBufferSize parameter when capturing changes in folders where you expect extensive changes. However, the larger the buffer size is, the less up-to-date the changes in the log file are, because updates are less frequent. A reasonable value is the average number of concurrent updates to the crawled folders.

```
<SESBufferSizeValue>
   <add sesBufferSize="n" />
</SESBufferSizeValue>
```

Installing the NTFS Web Service

Install this service after you install Oracle Search File Change Detector, as described in "Installing Oracle Search File Change Detector" on page 8-13.

To install the NTFS Web service on IIS 6 on Windows 2003 Server:

1. Copy `NTFSWebService.zip` from *ses_home*/search/lib/plugins/ntfs to the Windows server where Internet Information Services (IIS) is running.

2. Unzip the files in `NTFSWebService.zip` into a permanent folder.

3. Create a virtual directory on the Internet Information Server with the path pointing to the folder created in the previous step.

    a. Select **Administrative Tools** from the Windows Start menu, then select **Internet Information Services (IIS) Manager**.

    b. Expand the navigator in IIS Manager and right-click a Web site.

    c. Select **New**, then **Virtual Directory**.

    d. Follow the steps of the Virtual Directory Creation wizard.

4. On the Virtual Directory Access Permissions page of the wizard, select **Read** and **Run Scripts (such as ASP)**.

5. Open NTFSWebService Properties.

6. On the ASP.NET tab, verify that ASP.NET is version 2.0.

7. On the General tab, enter the settings described in Table 8–3.

8. On the Application tab, select Local Impersonation and enter the user credentials in the form *domain\username*.

   The application user must have these permissions:

   - Read on the NTFS Web Service physical directory

   - Read on the file share to be crawled.

   - Write on the C:\WINDOWS\Microsoft.NET\Framework*version*\Temporary ASP.NET Files folder.

     If the application user does not have access to this directory, then the Web service cannot load the required DLLs and signals the following error when it tries to access the Web service:

     ```
     Server Error in '/NTFSWS683343' Application
     Could not load file or assembly 'WEBSESNTFS' or one of its dependencies.
     Access is denied.
     ```

To install the NTFS Web service on IIS 7 on Windows 2008 Server:

1. Copy `NTFSWebService.zip` from *ses_home*/search/lib/plugins/ntfs to the Windows server where Internet Information Services (IIS) is running.

2. Unzip the files in `NTFSWebService.zip` to a directory.

3. Create a virtual directory on the Internet Information Server with the path pointing to the folder created in the previous step.

   a. Select **Administrative Tools** from the Windows Start menu, then select **roles > Web Server (IIS) > Internet Information Server**.

   b. Expand the navigator in IIS Manager, expand Sites, and right-click a Web site.

   c. Select **Add Virtual Diectory**.

   d. Follow the steps of the Virtual Directory Creation wizard.

4. Convert the virtual directory to application by right-clicking it, and then clicking **Convert to Application**.

5. Click the virtual directory, and in the rightmost pane, double click **Application Settings**.

6. Enter the settings described in Table 8–3.

7. Navigate to the directory where the files in `NTFSWebService.zip` were extracted as described in Step 2.

8. Open the `web.config` file in a text editor, and update the configuration settings in it, such that the application user must have the following permissions:

   - Read on the NTFS Web Service physical directory

   - Read on the file share to be crawled.

   - Write on the C:\WINDOWS\Microsoft.NET\Framework*version*\Temporary ASP.NET Files folder.

     If the application user does not have access to this directory, then the Web service cannot load the required DLLs and signals the following error when it tries to access the Web service:

     ```
     Server Error in '/NTFSWS683343' Application
     ```

```
Could not load file or assembly 'WEBSESNTFS' or one of its dependencies.
Access is denied.
```

Table 8–3   ASP.NET Configuration Settings

Parameter	Description
ServiceUsername	User name that authenticates Oracle SES to the NTFS Web service. You also enter this user name when creating the NTFS source. Oracle SES cannot access the Web service without the service username and password.
ServicePassword	Password for ServiceUsername. Ensure that this password is kept secure.
Batchsize	Determines the number of file URLs fetched for a Web service response. The NTFS connector processes a folder by fetching all the files in the folder.
FileChunkSize	Positive integer that specifies the chunk size. Large documents are sent in chunks to the NTFS connector. Enter a positive integer. For example, 1024000 divides the file into 1 MB chunks for sending over the Web.
	File chunk size should be the optimal data size that can transfer over the network.
IncrementalCrawlData	Path of the Results directory as specified in the Oracle Search File Change Detector configuration file. See "Modifying the File Change Detector Configuration File" on page 8-14.
	Choose the Application tab and impersonate as user that has read permission on the shared folder.

To verify that the NTFS Web service is installed correctly on IIS 6 on Windows 2003 Server:

1. Open Internet Information Services (IIS) Manager.

2. In the navigation tree, select **NTFSWebService** to display its contents in the right pane.

3. Right-click **NTFSWebService.asmx** and choose **Browse**.

4. Ensure that the Web service methods described in Table 8–4 are listed.

To verify that the NTFS Web service is installed correctly on IIS 7 on Windows 2008 Server:

1. Click the virtual directory that was created on the Internet Information Server as described in the Step 3 of section "To install the NTFS Web service on IIS 7 on Windows 2008 Server:" on page 8-17.

2. Click **Content View** at the bottom of the rightmost pane.

3. Right-click **NTFSWebService.asmx** and choose **Browse**.

4. Ensure that the Web service methods described in Table 8–4 are listed.

Table 8–4   NTFS Web Service Methods

Method	Description
ClearFCDLog	Clears the current Oracle Search File Change Detector log.
ClearPreviousFCDLog	Clears the previous Oracle Search File Change Detector log.
GetDFList	Gets all the files and subfolders in a specified folder.

Table 8–4   (Cont.) NTFS Web Service Methods

Method	Description
GetDocContainer	Gets the file and the access URL, display URL, and actual content after encoding. It also gets the ACL for the files and attributes of the file.
GetFileInParts	Gets the file after breaking it into chunk. The FileChunkSize parameter controls the chunk size.
GetMinimalMetadata	Fetches the ACL for the document and the last modified date of the file to determine whether the file has changed.
GetModifiedURLs	Gets a list of modified files and folders from the Oracle Search File Change Detector.

Configuring the NTFS Connector

The NTFS connector must be configured to perform incremental crawls with the Oracle Search File Change Detector. The connector has an additional parameter.

To configure the NTFS connector:

1. Open the Oracle SES Administration GUI, and select the **Sources** secondary tab.

2. Create or edit the NTFS connector.

3. Set the **Incremental crawl with the File Change Detector** parameter to `true`.

When the **Incremental crawl with File Change Detector** parameter is set to `true`, the NTFS connector performs the incremental crawl using the detector change logs. It reverts automatically to a scan-based incremental crawl under these conditions:

- The Oracle Search File Change Detector service is stopped.

- The Oracle Search File Change Detector service is started after the previous crawl start time. Scan-based incremental crawl is performed because some changes in the NTFS system might not be captured by the File Change Detector.

- The internal buffer of the File Change Detector overflowed. When the buffer overflows, the file change detector might not capture some changes.

To revert manually to a scan-based incremental crawl, set the **Incremental crawl with the File Change Detector** parameter to `false`.

Known Issues

- The Oracle Search File Change Detector does not capture changes to top-level directories used in the crawler configuration (UNC Path). Note that other directories within the folder are detected correctly.

- Changes to the source configuration, such as boundary rules and maximum file size, do not affect incremental crawls. For these changes to take effect, run a scan-based incremental crawl by setting the **Incremental crawl with the File Change Detector** parameter to false.

- File Change Detector hangs after the Windows Server Active Directory is restarted. You must manually restart the File Change Detector service whenever Active Directory is restarted.

Setting Up Oracle Calendar Sources

Oracle recommends creating one source group for *archived* calendar data and another source group for *active* calendar data. One instance for the archived source can run less

frequently, such as every week or month. This source should cover all history. A separate instance for the active source can run daily for only the most recent period.

Setting Up Identity Management for Oracle Calendar

The Oracle SES instance and the Oracle Calendar instance must be connected to the same Oracle Internet Directory system.

To set up a secure Oracle Calendar source:

1. On the Global Settings - Identity Management Setup page in the Oracle SES Administration GUI, select the **Oracle Internet Directory identity plug-in manager**, and click **Activate**.

2. Use the following LDIF file to create an **application entity** for the plug-in. An application entity is a data structure within LDAP used to represent and keep track of software applications accessing the directory with an LDAP client.

    oracle_home/bin/ldapmodify -h *oidHost* -p *OIDPortNumber* -D "cn=orcladmin" -w *password* -f *ses_home*/search/config/ldif/calPlugin.ldif

    This string defines the entity that is used for the plug-in:
    `orclapplicationcommonname=ocscalplugin,cn=oses,cn=products,cn=oraclecontext`. The entity has the password `welcome1`.

Creating an Oracle Calendar Source

Create an Oracle Calendar source on the Home - Sources page. Select Oracle Calendar from the Source Type list, and click **Create**. Enter values for the parameters described in Table 8–5.

Table 8–5   Calendar Source Parameters

Parameter	Value
Calendar server	http://*host name*:*port*
Application entity name	orclapplicationcommonname=ocscalplugin,cn=oses,cn=products,cn=oraclecontext
Application entity password	welcome1
OID server hostname	Oracle Internet Directory *hostname*
OID server port	389
OID server SSL port	636
OID server ldapbase	dc=us,dc=oracle,dc=com
OID login attribute	uid
User query	(objectclass=ctCalUser)
Past days	30
Future days	60
Rollover	true
Calendar server for Display URL	Calendar endpoint URL to be used to formulate the display URL; for example, http://calendarserver:7777.
	If this parameter is blank, then the value provided for the **Calendar server** parameter is used to formulate the display URL.

Oracle Calendar Attributes

- Description
- Priority
- Status
- start date
- end date
- event Type
- Author
- Created Date
- Title
- Location
- Dial_info
- ConferenceID
- ConferenceKey
- Duration

Setting Up Oracle Collaboration Suite E-Mail Sources

Oracle Collaboration Suite 10*g* Mail (Oracle Mail) implements the IMAP protocol, which is used by Oracle SES to retrieve data. You must login to the mail server using the user name and password to retrieve information. Note that Oracle Collaboration Suite mail server has a flag that allows the administrator to crawl mails of all users. The IMAP connector uses this feature to crawl all the mails of all the users using the mail server's admin login.

Important Notes for Oracle Collaboration Suite E-Mail Sources

Apart from the private folders, the Oracle Collaboration Suite E-Mail has shared folders. You can share any folder with another person by making it shared. Hence, while doing ACL stamping, the crawler must look if the mail is a part of a private folder or a shared folder and act accordingly.

The Oracle Collaboration Suite E-Mail has a Web interface to open mail. This same Web interface opens the searched mails from Oracle SES.

Required Tasks

For the e-mail admin to crawl data, set this parameter:

Go to Farm - Midtier - Mail Application - IMAP Server - Default Settings, and set **Allow Admin to Access Any Account** to true.

Setting Up Identity Management for Oracle Collaboration Suite E-Mail Sources

Activate the identity plug-in on the Global Settings - Identity Management Setup page. Select **Oracle Internet Directory** identity plug-in and click **Activate**.

Enter values for the following parameters:

- **Authentication Attribute**: Select **nickname**.

- **Host name**: Enter the host name of the computer where Oracle Internet Directory is running.

- **Port**: Enter the value `389`, which is the default LDAP port number.

- **Use SSL**: Enter `true` or `false`.

- **Realm**: Enter the Oracle Internet Directory realm; for example, `dc=us,dc=oracle,dc=com`.

- **User name**: Enter the Oracle Internet Directory administrator user name; for example, `cn=orcladmin`.

- **Password**: Enter the password for the user name.

Creating an Oracle Collaboration Suite E-Mail Sources

Create an Oracle Collaboration Suite E-Mail source on the Home - Sources page. Select **Oracle Collaboration Suite E-Mail** from the Source Type list, and click **Create**.

Enter values for the following parameters:

- **Email Server Address**: The IP address or DNS name of the IMAP e-mail server to be crawled, with the port number. This also specifies if the e-mail server follows IMAP or IMAPS protocol. Required.

  Use the format:

  `[imap | imaps]://IPaddress:portNumber`

  An exception is thrown if this parameter is null. If the server address is incorrect, then an exception is logged at the time of accessing the server.

- **Email Server Admin User**: The administration user name to access the e-mail server. Required.

- **Email Server Admin Password**: The password of the e-mail admin user. Required.

- **Remove Deleted messages from Index**: Indicates whether to keep the index for deleted mails in incremental recrawls. Valid values are `yes` and `no`. Any other value is considered to be `yes`.

- **Authentication Attribute**: Attribute used to validate the user. This varies based on the identity plug-in used for authentication. Oracle Collaboration Suite E-Mail uses Oracle Internet Directory for authentication, so set this parameter to `mail`.

- **LDAP Server**: The LDAP server information (IP address or DNS name, and so on).

- **LDAP Server Port**: The LDAP server port number.

- **LDAP Admin User Name**: The administrator user name of the LDAP server. Required.

- **LDAP Admin Password**: The password of the admin user of the LDAP server.

- **LDAP Base**: The domain to be searched; for example, `dc=oracle, dc=com`.

- **LDAP Query**: The query string defining the users whose e-mails must be crawled. This parameter is used for user-level partitioning.

  For example, to crawl only users with names beginning with A and having an e-mail in the domain us.example.com, the query is `(|(cn=A*)(mail=*@us.example.com))`.

- **Days from which crawling needs to be done**: The number of days in the past from which the crawling is done. The current date (time of crawl) is the base. For

example, a value of 200 specifies crawling messages with dates that are 200 or fewer days old. All mail is the default value.

- **Days to which the crawling needs to be done**: Specifies the number of days in the past to which the crawling must be done. The current date (time of crawl) is the base. For example, a value of 7 specifies crawling messages that are seven or more days old. Today is the default value.

- **Display URL template**: The display URL to be used for viewing the documents. This should have the placeholder for e-mail or user ID. For example, to see the full e-mail address in the display URL, enter the following:

```
http://<>/um/templates/message_list.uix?state=message_
list&cAction=openmessage&message_wmuid=$EMAIL
```

To see the user ID, enter the following:

```
http://<>/um/templates/message_list.uix?state=message_
list&cAction=openmessage&message_wmuid=$UID
```

- **Email Server Version**: The email server to be crawled. Valid values are `ocs10g` or `beehive`.

- **Folders to crawl**: The comma-delimited list of folders to be crawled. '*' means crawl all folders. Other valid values are INBOX, sent, and trash. This does not support regular expressions.

- **Folders not to crawl**: The comma-delimited list of folders not to be crawled. This list is considered only if the **Folders to crawl** parameter has the * wildcard as its value. Valid values are `INBOX`, `sent`, and `trash`. This parameter does not support regular expressions.

- **Revisit Skipped Attachments**: Controls whether the crawler revisits attachments that were skipped in earlier crawls because they did not meet the document type inclusion rules. This setting provides an alternative to a force recrawl after changing the document type inclusion rules. Set to `TRUE` to revisit skipped attachments, or set to `FALSE` otherwise (default).

- **Remove out-of-window messages**: Set to `true` to remove the crawled messages that are outside the current time window, else set to `false`. Default is `false`.

- **Incremental Crawl Mode**: Mode for the incremental crawl. The valid values are `New`, `Maintenance`, and `Auto`. Set to `New` to process only the new messages. Set to `Maintenance` to process only the deleted and moved messages. Set to `Auto` to let the crawler choose the right mode of incremental crawl for the crawling operation. Default is `New`.

- **Maintenance Time Window**: The timeframe to start `Maintenance` mode of incremental crawl in the `Auto` recrawl mode.

9

Configuring Access to Applications Sources

This chapter explains how to set up sources for Oracle and third-party databases and for Oracle business applications. It contains the following topics:

- Setting Up Database Sources
- Setting Up Oracle E-Business Suite Sources
- Setting Up Siebel 7.8 Sources
- Setting Up Siebel 8 Sources
- Setting Up PeopleSoft Sources
- Setting Up Oracle Fusion Sources
- Setting Up Oracle WebCenter Sources

Setting Up Database Sources

With a database source, you can crawl any JDBC-enabled database. A database source can crawl database content projected as a view or query. Each record in the view or query result set is interpreted as a document. You can create public database sources or secure database sources.

Required Columns in Database Sources

The view or query to be crawled must contain the columns described in Table 9–1. All column names must be in upper case.

Table 9–1   Database Source Required Columns

Column	Type	Description
CONTENT	VARCHAR2 or CLOB	Document content.
KEY	VARCHAR2	Key to identify the record in the record set. You can use a custom name for this column in the Administration GUI. See "Configuring the JDBC Driver" on page 9-3.
LANG	VARCHAR2	Document language in ISO 639-1 language code; for example, en for English or ja for Japanese.

Table 9–1   (Cont.) Database Source Required Columns

Column	Type	Description
LASTMODIFIEDDATE	DATE	Last modified date of the document.
		If you do not have a column for the mandatory LastModifiedDate attribute, use a constant date value in the SQL query for the source. Use a format that the getTimestamp method of the corresponding JDBC driver accepts without errors. Incremental changes to records are not picked up by re-crawls, so always schedule a full crawl.
URL	VARCHAR2	Display URL for the document. The value for this column cannot be null. This connector requires that there is URL-based access to the records in the result set of the view or query.

Optional Columns in Database Sources

The view or query can contain the optional columns describe in Table 9–2. Any other column is considered an attribute of the document.

If the query or view contains both content and either an attachment or attachment link, then one column (in the following order) is considered document content:

1. ATTACHMENT_LINK

2. ATTACHMENT

3. CONTENT

Even if the ATTACHMENT_LINK or ATTACHMENT column is specified in the query, you should include the mandatory CONTENT column. However, the content of ATTACHMENT_LINK or ATTACHMENT is indexed as document content.

Table 9–2   Database Source Optional Columns

Column	Type	Description
ATTACHMENT	BLOB	Binary attachments for the document.
ATTACHMENT_LINK	VARCHAR2	A link to the attachment for the document. HTTP, HTTPS, FILE, and FTP are valid.)
CONTENTTYPE	VARCHAR2	Content type of the document; for example, "text/html" for HTML documents, "application/pdf" for PDF documents, or "application/msword" for Microsoft Word documents.
		Leave blank when the content type is unknown or varied so that is it not feasible to specify the content type for each document individually.
PATH	VARCHAR2	Path to the document. This is used in the browse feature. This can be used to represent the organizational hierarchy of the document. For example, level1#level2#level3.
TITLE	VARCHAR2	Title of the document to be displayed in the Oracle SES search result page.

Table 9–2   (Cont.) Database Source Optional Columns

Column	Type	Description
LMD_TIMEZONE	VARCHAR2	Specifies the time zone for the date specified in LASTMODIFIEDDATE. For example CST. Oracle SES converts the last modified date from the specified time zone to Oracle SES time zone. If the time zone is not specified, then the date is considered to be in the Oracle SES time zone.

Configuring the JDBC Driver

Depending on your database source, you may need to configure the JDBC driver.

To crawl any third-party database:

1. Download the appropriate JDBC driver jar for JRE 1.6 into the *ses_home*/search/lib/plugins/oracleapplications directory.

2. Add the JRE 1.6 JDBC driver jar file name to the JDBC Driver Class parameter for a database type source in the Administration GUI, as described in Table 9–3.

3. Add the JRE 1.6 JDBC driver jar file name to the classpath in MANIFEST.MF of appsjdbc.jar and DBCrawler.jar.

4. Restart the middle tier.

For a key attribute that is not named KEY:

1. When configuring the database connector, specify the column name in the Key Attribute Name parameter, as described in Table 9–3.

2. In the crawling query, use the key attribute name as the alias for the key value column name. In this example, ID was entered as the value of the Key Attribute Name parameter and is the alias for KEYVAL:

```
SELECT keyval id, content, url, lastmodifieddate, lang FROM sales_only
```

Query File XML Schema Definition

The following is the XSD that defines the format of the XML query file.

```
<!--[if !supportEmptyParas]-->XSD for the XML sub-queries file:<!--[endif]-->
<?xml version="1.0" encoding="windows-1252" ?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns="http://xmlns.oracle.com/ses/sqlconnector/detail-attribute-queries"
targetNamespace="http://xmlns.oracle.com/ses/sqlconnector/detail-attribute-queries
"  elementFormDefault="qualified">
  <xsd:complexType name="sqlQueriesType">
    <xsd:annotation>
      <xsd:documentation>
        Specify detail and attribute queries as a source parameter for
        each document fetched by the parent query.
      </xsd:documentation>
    </xsd:annotation>
    <xsd:sequence>
      <xsd:element name="attachmentQueries" maxOccurs="1" minOccurs="0">
        <xsd:annotation>
          <xsd:documentation>
            Specify detail queries to fetch detail records for each document
            represented by the parent record. The parent records, fetched by
            the parent query, are specified as a source parameter. Each record
            in the document (parent) query can be associated with several detail
```

```
            (child) records. Each of these child records has a single column
            specifying the content that will be indexed as attachment to the
            parent document. The child query should select a single column, and
            the WHERE clause should have bind variables of the form
            ##PARENT ATTR##, where the value of PARENT ATTR from the parent
            record is substituted while executing the detail query.
          </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
          <xsd:sequence>
            <xsd:element name="query" maxOccurs="unbounded" minOccurs="1">
             <xsd:complexType>
                <!--Attribute to specify whether the contents retrieved by the
                    query is inline attachment or link to an attachment. The value
                    "true" specifies that the content is a link to an attachment
                    and "false" indicates inline attachment. Default value is
                    false.-->
                <xsd:attribute name="link" default="false"/>
                <!--Content type of the attachment. If no value is specified,
                    SES will auto-detect the content type.-->
                <xsd:attribute name="contenttype" default="null"/>
              </xsd:complexType>
            </xsd:element>
          </xsd:sequence>
        </xsd:complexType>
      </xsd:element>
      <xsd:element name="attributeQueries" maxOccurs="1" minOccurs="0">
        <xsd:annotation>
          <xsd:documentation>
            Specify queries to retrieve values of attributes of the parent
            document. Use this feature if the attribute can contain multiple
            values for a document. If the attribute is a single-valued
            attribute, then it can be specified in the parent query. The WHERE
            clause should have bind variables of the form ##PARENT ATTR##,
            where the value of PARENT ATTR from the parent record is substituted
            while executing the query.
          </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
          <xsd:sequence>
            <xsd:element name="query" maxOccurs="unbounded" minOccurs="1"/>
          </xsd:sequence>
        </xsd:complexType>
      </xsd:element>
    </xsd:sequence>
  </xsd:complexType>
  <xsd:element name="sqlQueries" type="sqlQueriesType"/>
</xsd:schema>
```

Creating Public Database Sources

Public database sources have no security implemented in Oracle SES.

To create a public database source:

1. Create a database source on the Home - Sources page. Select **Database** from the Source Type list, and click **Create**.

2. Enter the database source parameters as described in Table 9–3.

3. Click **Next**.

4. Set authorization to **No Access Control List**, and clear the authorization manager class name and jar name.

5. Click **Create** to create the database source.

Table 9–3   Database Source Parameters

Parameter	Value
Database Connection String	JDBC connection string for the database with content to be crawled. The JDBC string is driver-specific. For example, `jdbc:oracle:thin:@server:port:SID`
User ID	User ID to log in to the database specified in **Database Connection String**. This user ID must have access to the schema owning the view specified in **View** or the query specified in **Query**.
Password	Password to log in to the database specified in **Database Connection String**.
View	Table or view to be crawled. Specify either **View** or **Query**, not both.
JDBC Driver Class	JDBC driver class to connect to the database. For example, oracle.jdbc.driver.OracleDriver.
	Leave blank to use the default driver:
	■ Oracle Database: oracle.jdbc.driver.OracleDriver
	■ SQL Server: com.microsoft.sqlserver.jdbc.SQLServerDriver
Key Attribute Name	Name of the key column in the database source. The default value is KEY, as described in Table 9–1, " Database Source Required Columns".
Document Count	Maximum number of documents to be crawled before indexing. Enter -1 to crawl all documents before indexing.
Query File	Path to the XML file specifying the subqueries to crawl attachments and attributes of documents corresponding to every record in the main query. See "Query File XML Schema Definition" on page 9-3.
	When Oracle SES software is deployed across multiple middle tiers, then this query file must be accessible to all the middle tiers.
Query	Query projecting the content to be crawled. Specify either **View** or **Query**, not both.
URL Prefix	String that precedes the content of the URL column and forms a display URL for the document.
Cache File	Prefix of a local file name in which the contents can be temporarily cached while crawling.
Path Separator	The character separating the tokens in the PATH of the document as returned by the query or view. It must be a single character, and it cannot be a space, a single or double quote, or a control character.

Table 9–3   (Cont.) Database Source Parameters

Parameter	Value
Parse Attributes	Enter `true` to extract the values of the attributes from the document content specified in the `SOLUTION` or `CONTENT` column. Enter `false` otherwise, or when the content is type text/html.
	In this example, attr1 and attr2 are extracted as attributes of the document with values 22 and 333 respectively:
	`<attr1>22</attr1> <attr2>333</attr2>`
	Content up to the first attribute is interpreted as the document content. The remaining portion is used to extract attributes only. In this example, only "page" is considered document content:
	`page<attr1>22</attr1> is <attr2>333</attr2> dispersed`
Remove Deleted Documents	Enter `true` to remove deleted documents from the index; otherwise, enter `false`.
Attachment Link Authentication Type	Standard Java authentication type used by the application serving the link in the `ATTACHMENT_LINK` column. Enter one of these values:
	■ `PUBLIC`: No authentication.
	■ `DIGEST`: Digest authentication
	■ `BASIC`: Basic authentication
	■ `NATIVE`: Native authentication in the source
Attachment Link User ID	User ID for accessing the links specified in the `ATTACHMENT_LINK` column. Required when the link targets are secure.
Attachment Link Password	Password for **Attachment Link User ID**.
Attachment Link Realm	Realm of the application serving the link in the `ATTACHMENT_LINK` column. Required when the link targets are secure.
Grant Security Attributes	Space-separated list of grant security attributes. Leave blank for public sources.
Deny Security Attributes	Space-separated list of deny security attributes. Leave blank for public sources.
Oracle SSO Login URL	Set this parameter value when the authentication type is ORASSO. Oracle SES redirects the crawler to this SSO login URL for authentication before crawling the Database source.
	The format for this URL for Oracle 10g SSO secured Database source is:
	■ For basic authentication type:
	`https://`*server*`:`*port*`/pls/orasso/orasso.wwsso_app_ admin.ls_login`
	■ For form based authentication type:
	`https://`*server*`:`*port*`/mysso/signon.jsp`
	The format for this URL for Oracle 11g OAM secured Database source is:
	`http://`*server*`:`*port*`/oam/server/obrareq.cgi?encquery`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.

Table 9–3   (Cont.) Database Source Parameters

Parameter	Value
Oracle SSO Action URL	Set this parameter value when the authentication type is ORASSO. This is the URL that is displayed after successfully authenticating the SSO user before crawling the Database source.
	The format for this URL for Oracle 10g SSO secured Database source is:
	`https://server:port/sso/auth`
	The format for this URL for Oracle 11g OAM secured Database source is:
	`http://server:port/oam/server/auth_cred_submit`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.

Defining User-Defined Security for Database Sources

Some attributes in the view or query being crawled must be identified as security attributes. The values of these attributes determine if a user is authorized to view a document. These attributes can be either GRANT attributes or DENY attributes.

To create a database source with user-defined security:

1. On the **Home - Sources** page, select **Database** from the Source Type list and click **Create**.

2. Enter values for the parameters as described in Table 9–3. Specify the GRANT and DENY attributes as values for parameters **Grant Security Attributes** and **Deny Security Attributes** respectively. If there are multiple GRANT or DENY security attributes, then separate attribute names with a space.

3. Click **Next**.

4. Enter values for the authorization plug-in parameters:

   - **Authorization Database Connection String**: JDBC connection string for the authorization database. The values of the security attributes to which a given user is authorized are retrieved from this database. The JDBC string is driver-specific.

   - **User ID**: User ID to login to the authorization database.

   - **Password**: Password to login to the authorization database.

   - **Authorization Query**: SQL query to retrieve the values of security attributes to which a given user is authorized. The SELECT clause of this query should have all the security attributes specified in Step 2 with identical names. This query can be of two types:

     – The query can return a single record for a given user. The value in each security attribute column should be a space-delimited list of values to which the user is authorized.

     – The query can return multiple records for a given user. The value in each security attribute column of every row of the result set of this query is interpreted as a single value.

     Specify a question mark (?) as the placeholder for the username in the query.

- **Single Record Query**: Enter `true` if the authorization query returns a single record for a given user.

- **Authorization User ID Format**: Format of the user ID to be used in the SQL query specified in **Authorization Query**. This format should be an authentication attribute of the active identity plug-in.

  For example, if Oracle SES is configured with the Oracle Internet Directory identity plug-in (which supports DN, nickname and e-mail address as authentication attributes), then this parameter can be specified as nickname. The nickname of the current user is then used in the SQL authorization query to build the security filter.

  If no value is specified for this parameter, then the user ID in the canonical form of the active identity plug-in is used in the authorization query to build the security filter.

5. Click **Create** to create the database source.

Database Search Attributes

Database sources have no predefined attributes. The crawler collects attributes from columns defined during source creation. You must map the columns to the search attributes.

Example of Creating a Database Source With User-Defined Security

The document set to be crawled is in tables `T1` and `T2` as specified by the following query:

```
SELECT
     T1.ID,
     T1.DESCRIPTION,
     T2.NAME,
     T1.LAST_UPDATE_DATE,
     T2.AUTH_ID, T1.HIERARCHY
FROM
     T1, T2
WHERE
     T1.ID = T2.DOC_ID
```

The document content is provided by the `T1.DESCRIPTION` column.

Each document has an HTTP access URL of the form `http://my.company.com/docserver?doc_id=document_identifier`.

The value of `T2.AUTH_ID` controls access to a document. For example, user SCOTT can access a document only if the value of `T2.AUTH_ID` for the document is in the list of AUTH_IDs for SCOTT as retrieved by the following query:

```
SELECT AUTH_ID FROM USER_AUTH A
     WHERE A.USER='SCOTT'
```

This source can be crawled as a database source type with the following source parameter values:

- **Database Connection String**: `jdbc:oracle:thin:@example:7777:ses`
- **User ID**: `apps_user`
- **Password**: ******

- **View**:

- **Document Count**: -1

- **Query**:

```
SELECT
     'docserver?doc_id='|| T1.ID URL,
     T1.ID "KEY",
     'en' LANG,
     T1.LAST_UPDATE_DATE LASTMODIFIEDDATE,
     T1.DESCRIPTION CONTENT,
     'text/plain' CONTENTTYPE,
     T2.NAME CUSTOMER_NAME,
     T2.AUTH_ID,
     T1.HIERARCHY PATH
FROM
     T1, T2
WHERE
     T1.ID=T2.DOC_ID
```

- **Query File**:

```
<?xml version="1.0" encoding="UTF-8" ?>
<sqlQueries xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://xmlns.oracle.com/ses/sqlconnector/detail-attribute-q
ueries detail-attribute-queries.xsd"
 xmlns="http://xmlns.oracle.com/ses/sqlconnector/detail-attribute-queries">
   <attachmentQueries>
     <query>
       <![CDATA[SELECT COMMENTS FROM COMMENTS_TBL WHERE DOCID=##KEY##]]>
     </query>
     <query link="true" contenttype='text/html'>
           <![CDATA[SELECT NOTES FROM NOTES_LINK_TBL WHERE DOCID=##KEY##]]>
     </query>
   </attachmentQueries>
   <attributeQueries>
       <query>
         <![CDATA[SELECT AUTHOR FROM AUTHORS_TBL WHERE DOCID=##KEY##]]>
       </query>
       <query>
               <![CDATA[SELECT KEYWORD FROM KEYWORD_TBL WHERE DOCID=##KEY##]]>
       </query>
   </attributeQueries>
</sqlQueries>
```

AUTHID and KEY are columns in the select list of the parent query.

Note: This must be the path to a local file containing the subqueries for attributes and attachments that are currently listed directly for query file.

- **URL Prefix**: http://my.example.com/

- **Cache File**: /tmp/cacheFile

- **Path Separator**: #

- **Parse Attributes**: false

- **Grant Security Attributes**: AUTH_ID

- **Deny Security Attributes**:

The following are sample parameter values for authorization.

- **Database Connection String**: `jdbc:oracle:thin:@example:7777:ses`

- **User ID**: `apps_user`

- **Password**: *****

- **Authorization Query**:

  ```
  SELECT AUTH_ID
  FROM USER_AUTH A
  WHERE A.USER=UPPER(?)
  ```

- **Single Record Query**: `false`

- **Authorization User ID Format**: `username`

Setting Up Oracle E-Business Suite Sources

The Oracle E-Business Suite connector uses the Oracle SES XML connector framework, where searching is based on Oracle E-Business Suite data available as XML feeds.

> **See Also:** "Overview of XML Connector Framework" on page 4-11

To activate an identity plug-in for Oracle E-Business Suite sources:

1. On the Global Settings page, select **Identity Management Setup**.

2. Select **Oracle E-Business Suite** and click **Activate** to display the Activate Identity Plug-in page.

3. Enter values for the parameters as described in Table 9–4. Obtain the values for these parameters from the E-Business Suite administrator.

4. Click **Finish**.

Table 9–4   Oracle E-Business Suite Identity Management Parameters

Parameter	Value
HTTP endpoint for authentication	HTTP endpoint of Oracle E-Business Suite that provides the user authentication and validation service.
User ID	Administrator user ID for posting data to the endpoint specified in **HTTP endpoint for authentication**.
Password	Password for **User ID**.

To create an Oracle E-Business Suite source:

1. Activate an identity plug-in as described in the previous procedure.

2. On the home page, select the **Sources** secondary tab.

3. Select **Oracle E-Business Suite** from the Source Type list, and click **Create**.

4. Enter the source parameters as described in Table 9–5.

5. Click **Next**.

6. Click **Get Parameters** to obtain a list of parameters for the authorization manager plug-in.

7. Enter the values for the authorization manager plug-in parameters as described in Table 9–6.

8. Click **Create**.

After processing each data feed, the crawler uploads a status feed to the location specified in the XML configuration file specified in the Configuration URL parameter. This status feed has a name in the following format:

- *datafeedFilename*.suc when the data feed was processed successfully.

- *datafeedFilename*.err when an error occurred during processing. The errors are listed in this file.

Table 9–5   Oracle E-Business Suite Source Parameters

Parameter	Value
Configuration URL	URL of the XML configuration file providing details of the source, such as the data feed type, location, security attributes, and so on. The URL is a HTTP URL accessible over HTTP. Obtain this file from the Oracle E-Business Suite administrator.
Authentication Type	Enter the value Native.
User ID	User ID to access the data feeds. The access details of the data feed are specified in the configuration file. The user id can be obtained from Oracle E-Business Suite administrator.
Password	Password for **User ID**.
Realm	The realm of the application serving the feeds. The parameter is usually left blank.

Oracle SSO Login URL	Set this parameter value when the authentication type is ORASSO. Oracle SES redirects the crawler to this SSO login URL for authentication before crawling the Oracle E-Business Suite source.
	
	The format for this URL for Oracle 10g SSO secured Oracle E-Business Suite source is:
	
	  ■ For basic authentication type:
	
	    `https://server:port/pls/orasso/orasso.wwsso_app_admin.ls_login`
	
	  ■ For form based authentication type:
	
	    `https://server:port/mysso/signon.jsp`
	
	The format for this URL for Oracle 11g OAM secured Oracle E-Business Suite source is:
	
	`http://server:port/oam/server/obrareq.cgi?encquery`
	
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.

Oracle SSO Action URL	Set this parameter value when the authentication type is ORASSO. This is the URL that is displayed after successfully authenticating the SSO user before crawling the Oracle E-Business Suite source.
	
	The format for this URL for Oracle 10g SSO secured Oracle E-Business Suite source is:
	
	`https://server:port/sso/auth`
	
	The format for this URL for Oracle 11g OAM secured Oracle E-Business Suite source is:
	
	`http://server:port/oam/server/auth_cred_submit`
	
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.

Scratch Directory	A directory on the same computer as Oracle SES, where the status logs are created temporarily.

Table 9–5 (Cont.) Oracle E-Business Suite Source Parameters

Parameter	Value
Maximum number of connection attempts	Maximum number of attempts to connect to the target server to access the data feed.
Delete Linked Documents	Set it to `true` if the documents crawled from the links in the feeds must be deleted. Set it to `false` otherwise. Default is `true`.
Number of data feeds to be pre-fetched	Number of data feeds to pre-fetch. Its value can be any number greater than or equal to 0. If its value is set to 0, then no data feed is pre-fetched. The value specified must be such that there is sufficient memory to cache the pre-fetched feeds. The default value is 0.
Enable Resume Crawl	Set it to `true` to enable resuming crawl when the crawl fails or stops. The default value is `true`.
Stop Crawl On Content Fetch Error	Set it to `true` to stop crawl on content fetch error, else set it to `false`. The default value is `false`.

Table 9–6 Oracle E-Business Suite Authorization Parameters

Parameter	Value
HTTP endpoint for authorization	HTTP endpoint of E-Business Suite that provides the user authorization service.
User ID	User ID.
Password	Password for **User ID**.
Business Component	Name of the Oracle E-Business Suite business component being crawled. The values of the security attributes for which the current user is authorized in the realm of this business component is retrieved to build the security filter for the user when the user logs into Oracle SES. For example, `oracle.apps.fnd.fwk.search.NavigationSVO`.
Security attribute values for anonymous user	Comma-delimited list of authorized values of security attributes for anonymous user. If the parameter is left blank, then the authorization service is contacted to retrieve the values of security attributes accessible for anonymous user.
Display URL Prefix	HTTP host information to prefix the partial URL specified in the access URL of the documents in XML feeds to form the complete URL. This complete URL is the display URL of the document when the document link in the Oracle SES search results page is clicked.
	This value must form a valid URL when concatenated with the access URL element of an item in the data feed. Be careful to avoid having either two slashes or none when the values are combined. Thus, enter a trailing slash (/) if the access URLs do not begin with a slash, or omit the trailing slash from the prefix if the access URLs begin with a slash.

Setting Up Siebel 7.8 Sources

The Siebel 7.8 source crawler is based on crawling a view or query in a database. Each record in the view or query is considered a document.

The Siebel 7.8 connector supports Siebel installations on the following databases:

- Oracle
- SQL Server

Requirements for Siebel 7.8 Sources

Views and queries to be crawled must contain the columns described in Table 9–7.

Table 9–7   Siebel 7.8 Source Required Columns

Column Name	Datatype	Description
URL	VARCHAR2	Display URL for the document. The value for this column cannot be null. This connector requires that there is URL-based access to the records in the result set of the view or query.
SOLUTION or CONTENT	VARCHAR2 or CLOB	Document content.
LASTMODIFIEDDATE	DATE	Last modified date for crawl.
KEY	VARCHAR2	Primary key of the records.
LANG	VARCHAR2	Document language, such as en for English or ja for Japanese.

Any other columns in the views or queries is considered an attribute of the document.

Installing the JDBC Driver for Microsoft SQL Server

If Siebel 7.8 is installed over a Microsoft SQL Server database, then Oracle SES must have access to the JDBC driver for SQL Server.

To install the SQL Server JDBC driver for Oracle SES:

1. Download the SQL Server 2005 JDBC driver 2.0 from http://www.microsoft.com/download/en/details.aspx?id=2505.

2. Follow the instructions at the same location to install the driver.

3. Copy sqljdbc4.jar for JRE 1.6 from the installed directory to *ses_home*/search/lib/plugins/oracleapplications/.

4. Add sqljdbc4.jar to the classpath in MANIFEST.MF of appsjdbc.jar and DBCrawler.jar.

5. Restart the middle tier.

Enabling Crawling of Attachments for Siebel 7.8

The Oracle SES crawler can be used to crawl Siebel 7.8 attachments on remote computers using remote method invocation (RMI). To use RMI, you must set up:

- The RMI server on the Siebel host to decompress Siebel attachment files.

- A source on the SES host to crawl Siebel 7.8 attachments over RMI.

> **See Also:** "Starting the Decompression Server" on page 9-13

Starting the Decompression Server

If you are using Windows Native Access for the decompression server, then you can skip this procedure.

To start the decompression server on the Siebel host:

1. Install JDK if it is not already installed on the Siebel host.

2. Set the JAVA_HOME environment variable to the JDK folder.

3. Update the PATH environment variable to begin with `JAVA_HOME\bin`.

4. Start the RMI registry:

   - To use default port 1099, double-click `JAVA_HOME\bin\rmiregistry.exe`.

   - To use a different port number, execute this command from the Windows command prompt, replacing *port_number* as desired:

     `%JAVA_HOME%\bin\rmiregistry` *port_number*

5. Copy this file from the Oracle SES host to the Siebel host and unzip it into the `SiebelDecompSvr` folder:

   ses_home`/search/lib/plugins/oracleapplications/DecompressionServer.zip`

6. Update `SiebelDecompSvr\startDecompServer.bat` by replacing the placeholders in the `java` command with the following information:

   - Absolute path to the `sseunzip.exe` decompression utility.

   - Folder where Siebel attachment files can be temporarily decompressed.

   - Level for logging decompression server output: FINE, INFO, WARNING, SEVERE. Default is INFO.

   - For user credential-based security, the user name specified in the Oracle SES configuration. See Table 9–8, " Siebel 7.8 Identity Management Parameters".

   - For user credential-based security, the password specified in the Oracle SES configuration.

7. For RMI security, replace the placeholders in the `sec.policy` security policy file with the IP address or host name of the Oracle SES instances that crawls the Siebel attachments.

8. Double-click `SiebelDecompSvr\startDecompServer.bat` on the Siebel host to start the decompression server.

Setting Up Identity Management for Siebel 7.8

This procedure enables Oracle SES to validate users by querying the Siebel 7.8 identity management system.

To activate the Siebel 7.8 identity plug-in:

1. On the Global Settings page, select Identity Management Setup under the System heading.

   The **Global Settings - Identity Management Setup** page is displayed.

2. Select **Siebel 7.8** and click **Activate**.

3. Enter values for the parameters described in Table 9–8, then click **Finish**.

Table 9–8   Siebel 7.8 Identity Management Parameters

Parameter	Value
Authentication and Validation Database Connection String	JDBC connection string for the Siebel 7.8 database for authenticating and validating users.
User ID	Administrator ID for the Siebel 7.8 database (specified in **Authentication and Validation Database Connection String**) for validating users.

Table 9–8 (Cont.) Siebel 7.8 Identity Management Parameters

Parameter	Value
Password	Password for **User ID**.
User Validation Query	SQL query for validating users. The query must return 1 if the user is valid, and null otherwise. Use a question mark (?) as a placeholder for the user name. This query replaces this default query: `SELECT 1 FROM dbo.S_USER WHERE LOGIN=upper(?)`

Creating a Secured Siebel 7.8 Source

Oracle SES supports these Siebel 7.8 secured business components: Activity, Accounts, Contacts, Literature, Products, and Service Request.

To create a source for Siebel 7.8 secured business components:

1. On the home page, click the **Sources** secondary tab to display the Sources page.

2. Select **Siebel 7.8** from the **Source Type** list, then click **Create** to display Step 1 Parameters.

3. Complete the form, entering values for the parameters described in Table 9–9.

4. Click **Next** to display Step 2 Authorization.

5. Provide values for the authorization parameters described in Table 9–10.

6. Click **Create**.

Table 9–9 Siebel 7.8 Source Parameters (Step 1)

Parameter	Description
Database Connection String	JDBC connection string for the Siebel 7.8 database from which the content has to be crawled.
User ID	User ID to login to the Siebel 7.8 database specified in **Database Connection String**. This user ID should have access to the schema owning the view specified in **View** or the query specified in **Query**.
Password	Password to login to the Siebel 7.8 database specified in **Database Connection String**.
View	Table or view with the columns needed for crawling. Leave this parameter blank.
Document Count	Maximum number of documents to be crawled before indexing. Enter -1 for this parameter.
Query	Query projecting the columns for crawling. Specify the query for the required business component given in "Queries to Crawl Siebel 7.8 Business Components" on page 9-17.
Query File	Path to the XML file specifying the subqueries to crawl attachments and attributes of documents corresponding to every record in the main query. Leave this parameter blank.
URL Prefix	String to prefix the content of URL column to form a complete display URL for the document
Cache File	Local file to which the contents can be temporarily cached while crawling.

Table 9–9   (Cont.) Siebel 7.8 Source Parameters (Step 1)

Parameter	Description
Path Separator	Path separator character in the document path string.
	Leave this parameter blank.
Parse Attributes	Extracts attribute values from the document content specified in the SOLUTION or CONTENT column.
	Enter `false` for this parameter.
Grant Security Attributes	Space-separated list of grant security attributes.
	Enter `VISIBILITYID` for this parameter.
Deny Security Attributes	Space-separated list of deny security attributes.
	Leave this parameter blank.
Remove Deleted Documents	Set this parameter to `false` to optimize incremental crawl. Only records that have changed since the previous crawl are crawled. The documents deleted in the database are not deleted from the Oracle SES index.
	Set this parameter to `true` if records that have been deleted from the database since the previous full crawl, or since a previous incremental crawl with this parameter set to `true`, should be deleted from the Oracle SES index as well.
	Crawling may take longer when this parameter is set to `true` than when it is set to `false`.
Attachment Link Authentication Type	Enter `NATIVE` for user credential-based security; otherwise, enter `PUBLIC` for this parameter.
Attachment Link User ID	User ID for accessing the link in the attachment link column.
	If the Attachment Link Authentication Type is `NATIVE` and attachments are accessed using RMI, then enter the user name provided in `SiebelDecompSvr\startDecompServer.bat`; otherwise, leave this parameter blank.
Attachment Link Password	Password for accessing the link in the attachment link column.
	If the Attachment Link Authentication Type is `NATIVE` and Attachments are accessed using RMI, then enter the password provided in `SiebelDecompSvr\startDecompServer.bat`; otherwise, leave this parameter blank.
Attachment Link Realm	Realm of the application serving the link in the attachment link column.
	Leave this parameter blank.

Table 9–10   Siebel 7.8 Authorization Parameters (Step 2)

Parameter	Value
Authorization Database Connection String	JDBC connection string for the authorization database. The values of the visibility IDs for users are retrieved from this database. Typically, this connection string matches the **Authentication and Validation Database Connection String** on Page 1.
User ID	Administrator user ID for the authorization database specified in **Authorization Database Connection String**
Password	Password for **User ID**.
	Note: Oracle SES does not allow a password to contain the characters `&`, `\`, `"`, `<`, or `>`.

Table 9–10   (Cont.) Siebel 7.8 Authorization Parameters (Step 2)

Parameter	Value
Authorization Query	SQL query to retrieve the values of visibility IDs. Use a question mark (?) as a placeholder for the user name. See Example 9–1 for more information about authorization queries.

Example 9–1   Siebel 7.8 Authorization Queries

The following query is the default authorization query that can be used for the Service Request, Accounts, Products, Literature, Solution, and Contacts business components:

```
SELECT p.BU_ID visibilityid
FROM dbo.S_POSTN p
    INNER JOIN dbo.S_CONTACT c2 ON c2.PR_HELD_POSTN_ID = p.ROW_ID
    INNER JOIN dbo.S_USER    u  ON u.PAR_ROW_ID        = c2.PAR_ROW_ID
WHERE u.LOGIN = upper(?)
```

Use the following authorization query for Activity business components:

```
SELECT DISTINCT usr.LOGIN visibilityid
FROM
    S_PARTY_RPT_REL rpt,
    S_POSTN postn,
    S_USER usr,
    S_PARTY pty
WHERE
    rpt.PARTY_ID IN (
        SELECT T3.PR_HELD_POSTN_ID
        FROM dbo.S_PARTY T1
            INNER JOIN dbo.S_EMP_PER T2 ON T1.ROW_ID = T2.PAR_ROW_ID
            INNER JOIN dbo.S_CONTACT T3 ON T1.ROW_ID = T3.PAR_ROW_ID
            INNER JOIN dbo.S_USER    T4 ON T1.ROW_ID = T4.PAR_ROW_ID
        WHERE (T3.EMP_FLG = 'Y') AND T4.LOGIN=upper(?))
    AND rpt.SUB_PARTY_ID = postn.PAR_ROW_ID
    AND postn.PR_EMP_ID = usr.ROW_ID
    AND usr.PAR_ROW_ID = pty.ROW_ID
```

See "Queries to Crawl Siebel 7.8 Business Components" on page 9-17.

Creating a Public Siebel 7.8 Source

Oracle SES supports Solution as a public business component.

To create a source for Siebel 7.8 public business components:

1. On the home page, click the **Sources** secondary tab to display the Sources page.

2. Select **Siebel 7.8 (Public)** from the **Source Type** list, then click **Create**.

3. Complete the form, entering values for the parameters described in Table 9–9. The two parameters to grant and deny security attributes are omitted from the configuration of a public source.

4. Click **Create**.

Queries to Crawl Siebel 7.8 Business Components

This section includes the queries to crawl the Siebel 7.8 business components supported by Oracle SES:

■   Service Request Attachments

- [Accounts](#)

- [Products](#)

- [Literature](#)

- [Solution](#)

- [Service Request](#)

- [Contacts](#)

- [Activity](#)

- [Activity Attachment](#)

> **Note:** The Siebel 7.8 crawling queries provided in this section are supported only when used with default Siebel 7.8 database schema. Neither these queries nor any modifications to these queries are supported in any modified Siebel 7.8 database schema.

> **Note:** Set the "Parse Attributes" parameter value of the Siebel 7.8 data source to `false` before executing these queries.

The queries appear in two forms: A multi-line query for readability and a single-line version to cut-and-paste into the Query parameter. Queries use SQL Server syntax unless otherwise noted and must be modified slightly for use with Oracle Database.

To use a sample query as the value of the Query parameter:

- Replace *HostName* with the name of the host where Siebel is installed.

- The values of the parameters `SWEView` and `SWEApplet0` in the queries are the names of views and applets in a default Siebel installation. Change them as required if different names were used while installing Siebel 7.8.

- Add appropriate `WHERE` clauses to these queries depending on the search specification of views, applets and business components in the Siebel system. For example, if the Siebel system is configured to locate only internal service requests, then append the `WHERE` clause to the query for Service Request business component as follows: `WHERE c.SR_TYPE_CD = 'Internal'`.

To modify a query for use in Oracle Database instead of SQL Server:

- Replace the string concatenation operator '+' with '||'.

- Replace the table owner name `dbo` with the appropriate table owner name in Oracle Database.

Service Request Attachments

```
SELECT
    'callcenter_enu/start.swe?SWECmd=GotoView
        &SWEView=All+Service+Request+across+Organizations
        &SWERF=1
        &SWEHo=HostName
        &SWEBU=1
        &SWEApplet0=Service+Request+Detail+Applet
        &SWERowId0='+c.PAR_ROW_ID+'
        &SRAttId='+c.ROW_ID URL,
    'US' LANG,
```

```
        c.LAST_UPD LASTMODIFIEDDATE,
    'text/html' CONTENTTYPE,
    c.ROW_ID "KEY",
    coalesce('<b>Attachment Name:</b> '+c.FILE_NAME,'<null>')
        +coalesce(',<br><b>SR Number:</b> '+srv.SR_NUM,'<null>')
        +coalesce(',<br><b>SR Summary: </b>'+srv.SR_TITLE,'<null>') SOLUTION,
    c.ROW_ID sblrowid,
    c.CREATED created_on,
    c.CREATED_BY createdby,
    c.LAST_UPD_BY lastupdatedby,
    c.PAR_ROW_ID title,
    c.FILE_SRC_TYPE "type",
    c.FILE_EXT code01,
    c.COMMENTS "comment",
    c.FILE_SRC_PATH location,
    'Service Request Attachment' sblbctype,
    usr.LOGIN owner,
    srv.BU_ID visibilityid
FROM
    dbo.S_SR_ATT c
    INNER JOIN dbo.S_SRV_REQ srv      ON c.PAR_ROW_ID   = srv.ROW_ID
    LEFT OUTER JOIN dbo.S_USER usr    ON usr.PAR_ROW_ID = srv.OWNER_EMP_ID
    LEFT OUTER JOIN dbo.S_CONTACT con ON con.PAR_ROW_ID = c.LAST_UPD_BY
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT 'callcenter_
enu/start.swe?SWECmd=GotoView&SWEView=All+Service+Request+across+Organizations&SWE
RF=1&SWEHo=HostName&SWEBU=1&SWEApplet0=Service+Request+Detail+Applet&SWERowId0='+c
.PAR_ROW_ID+'&SRAttId='+c.ROW_ID URL, 'US' LANG, c.LAST_UPD LASTMODIFIEDDATE,
'text/html' CONTENTTYPE, c.ROW_ID "KEY", coalesce('<b>Attachment Name:</b>
'+c.FILE_NAME,'<null>')+coalesce(',<br><b>SR Number:</b> '+srv.SR_
NUM,'<null>')+coalesce(',<br><b>SR Summary: </b>'+srv.SR_TITLE,'<null>') SOLUTION,
c.ROW_ID sblrowid, c.CREATED created_on, c.CREATED_BY createdby, c.LAST_UPD_BY
lastupdatedby, c.PAR_ROW_ID title, c.FILE_SRC_TYPE "type",c.FILE_EXT code01,
c.COMMENTS "comment", c.FILE_SRC_PATH location, 'Service Request Attachment'
sblbctype, usr.LOGIN owner, srv.BU_ID visibilityid FROM dbo.S_SR_ATT c INNER JOIN
dbo.S_SRV_REQ srv ON c.PAR_ROW_ID=srv.ROW_ID LEFT OUTER JOIN dbo.S_USER usr ON
usr.PAR_ROW_ID = srv.OWNER_EMP_ID LEFT OUTER JOIN dbo.S_CONTACT con ON con.PAR_
ROW_ID = c.LAST_UPD_BY
```

Accounts

```
SELECT
    'callcenter_enu/start.swe?SWECmd=GotoView
        &SWEView=All+Accounts+across+Organizations
        &SWERF=1
        &SWEHo=HostName
        &SWEBU=1
        &SWEApplet0=Account+List+Applet
        &SWERowId0='+T1.ROW_ID URL,
    'US' LANG,
    T2.LAST_UPD LASTMODIFIEDDATE,
    'text/html' CONTENTTYPE,
    T1.ROW_ID "KEY",
    coalesce('<b>Name:</b> '+T2.NAME,'<null>')
        +coalesce(',<br><b>Type:</b> '+T2.OU_TYPE_CD,'<null>')+',<br>
            <b>Address:</b>'
        +coalesce(T5.ADDR,'<null>')
        +coalesce(','+T5.CITY,'<null>')
```

```
         +coalesce(','+T5.STATE+' '+T5.ZIPCODE,'<null>')
         +coalesce(','+T5.COUNTRY,'<null>') SOLUTION,
      T1.ROW_ID sblrowid,
      T2.CREATED created_on,
      T2.CREATED_BY createdby,
      T2.LAST_UPD_BY lastupdatedby,
      T2.NAME title,
      T2.OU_NUM csn,
      T2.OU_TYPE_CD type,
      T2.LOC location,
      T10.LOGIN alias,
      T5.ADDR street,
      T5.CITY city,
      T5.STATE state,
      T5.COUNTRY country,
      T5.ZIPCODE zipcode,
      'Account' sblbctype,
      T2.BU_ID visibilityid
FROM
      dbo.S_PARTY T1
      INNER JOIN dbo.S_ORG_EXT T2        ON T1.ROW_ID      = T2.PAR_ROW_ID
      INNER JOIN dbo.S_ACCNT_POSTN T3    ON T2.PR_POSTN_ID = T3.POSITION_ID
         AND T2.ROW_ID = T3.OU_EXT_ID
      INNER JOIN dbo.S_PARTY T4          ON T3.POSITION_ID = T4.ROW_ID
      LEFT OUTER JOIN dbo.S_POSTN T9     ON T3.POSITION_ID = T9.PAR_ROW_ID
      LEFT OUTER JOIN dbo.S_ADDR_ORG T5  ON T2.PR_ADDR_ID  = T5.ROW_ID
      LEFT OUTER JOIN dbo.S_USER T10     ON T9.PR_EMP_ID   = T10.PAR_ROW_ID
      LEFT OUTER JOIN dbo.S_CONTACT T11  ON T11.PAR_ROW_ID = T2.LAST_UPD_BY
WHERE
      (T2.INT_ORG_FLG != 'Y' OR T2.PRTNR_FLG = 'Y')
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT 'callcenter_
enu/start.swe?SWECmd=GotoView&SWEView=All+Accounts+across+Organizations&SWERF=1&SW
EHo=HostName&SWEBU=1&SWEApplet0=Account+List+Applet&SWERowId0='+T1.ROW_ID URL,
'US' LANG, T2.LAST_UPD LASTMODIFIEDDATE, 'text/html' CONTENTTYPE, T1.ROW_ID "KEY",
coalesce('<b>Name:</b> '+T2.NAME,'<null>')+coalesce(',<br><b>Type:</b> '+T2.OU_
TYPE_CD,'<null>')+',<br><b>Address:</b>
'+coalesce(T5.ADDR,'<null>')+coalesce(','+T5.CITY,'<null>')+coalesce(','+T5.STATE+
' '+T5.ZIPCODE,'<null>')+coalesce(','+T5.COUNTRY,'<null>') SOLUTION, T1.ROW_
ID sblrowid, T2.CREATED created_on, T2.CREATED_BY createdby, T2.LAST_UPD_BY
lastupdatedby, T2.NAME title, T2.OU_NUM csn, T2.OU_TYPE_CD type, T2.LOC location,
T10.LOGIN alias, T5.ADDR street, T5.CITY city, T5.STATE state, T5.COUNTRY country,
T5.ZIPCODE zipcode, 'Account' sblbctype, T2.BU_ID visibilityid FROM dbo.S_PARTY T1
INNER JOIN dbo.S_ORG_EXT T2 ON T1.ROW_ID = T2.PAR_ROW_ID INNER JOIN dbo.S_ACCNT_
POSTN T3 ON T2.PR_POSTN_ID = T3.POSITION_ID AND T2.ROW_ID = T3.OU_EXT_ID INNER
JOIN dbo.S_PARTY T4 ON T3.POSITION_ID = T4.ROW_ID LEFT OUTER JOIN dbo.S_POSTN T9
ON T3.POSITION_ID = T9.PAR_ROW_ID LEFT OUTER JOIN dbo.S_ADDR_ORG T5 ON T2.PR_ADDR_
ID=T5.ROW_ID LEFT OUTER JOIN dbo.S_USER T10 ON T9.PR_EMP_ID = T10.PAR_ROW_ID LEFT
OUTER JOIN dbo.S_CONTACT T11 ON T11.PAR_ROW_ID=T2.LAST_UPD_BY WHERE (T2.INT_ORG_
FLG != 'Y' OR T2.PRTNR_FLG = 'Y')
```

Products

```
SELECT
      'callcenter_enu/start.swe?SWECmd=GotoView
         &SWEView=All+Products+across+Organizations
         &SWERF=1
         &SWEHo=HostName
```

```
        &SWEBU=1
        &SWEApplet0=Product+List+Applet
        &SWERowId0='+c.ROW_ID URL,
    'US' LANG,
    c.LAST_UPD LASTMODIFIEDDATE,
    'text/html' CONTENTTYPE,
    c.ROW_ID "KEY",
    coalesce('<b>Name:</b> '+ c.NAME,'<null>')
        +coalesce(',<br><b>Part Number:</b> '+c.VENDR_PART_NUM,'<null>')
        +coalesce(',<br><b>Catalog/Category:</b> '+ c2.NAME,'<null>') SOLUTION,
    c.DESC_TEXT description,
    c.ROW_ID sblrowid,
    c.CREATED created_on,
    c.CREATED_BY createdby,
    c.NAME title,
    'Product Catalog' sblbctype,
    c.VENDR_PART_NUM name,
    c.VENDR_PART_NUM + ' ' + c3.PROD_ID + ' ' + c3.CTLG_CAT_ID summary,
    c.BU_ID visibilityid,
    c2.NAME sblvisibilityinfo,
    c.VERSION type
FROM
    dbo.S_PROD_INT c
    INNER JOIN      dbo.S_CTLG_CAT_PROD c3 ON c3.PROD_ID    = c.ROW_ID
    INNER JOIN      dbo.S_CTLG_CAT      c2 ON c2.ROW_ID     = c3.CTLG_CAT_ID
    LEFT OUTER JOIN dbo.S_CONTACT       c4 ON c4.PAR_ROW_ID = c.LAST_UPD_BY
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT 'callcenter_
enu/start.swe?SWECmd=GotoView&SWEView=All+Products+across+Organizations&SWERF=1&SW
EHo=HostName&SWEBU=1&SWEApplet0=Product+List+Applet&SWERowId0='+c.ROW_ID URL, 'US'
LANG, c.LAST_UPD LASTMODIFIEDDATE, 'text/html' CONTENTTYPE, c.ROW_ID "KEY",
coalesce('<b>Name:</b> '+ c.NAME,'<null>')+coalesce(',<br><b>Part Number:</b>
'+c.VENDR_PART_NUM,'<null>')+coalesce(',<br><b>Catalog/Category:</b> '+
c2.NAME,'<null>') SOLUTION, c.DESC_TEXT description, c.ROW_ID sblrowid, c.CREATED
created_on, c.CREATED_BY createdby, c.NAME title, 'Product Catalog' sblbctype,
c.VENDR_PART_NUM name, c.VENDR_PART_NUM + ' ' + c3.PROD_ID + ' ' + c3.CTLG_CAT_ID
summary, c.BU_ID visibilityid, c2.NAME sblvisibilityinfo, c.VERSION type FROM
dbo.S_PROD_INT c INNER JOIN dbo.S_CTLG_CAT_PROD c3 ON c3.PROD_ID=c.ROW_ID INNER
JOIN dbo.S_CTLG_CAT c2 ON c2.ROW_ID=c3.CTLG_CAT_ID LEFT OUTER JOIN dbo.S_CONTACT
c4 ON c4.PAR_ROW_ID=c.LAST_UPD_BY
```

Literature

```
SELECT
    'callcenter_enu/start.swe?SWECmd=GotoView
        &SWEView=All+Sales+Tools+across+Organizations
        &SWERF=1
        &SWEHo=HostName
        &SWEBU=1
        &SWEApplet0=Sales+Tool+List+Applet
        &SWERowId0='+c.ROW_ID URL,
    'US' LANG,
    c.LAST_UPD LASTMODIFIEDDATE,
    'text/html' CONTENTTYPE,
    c.LAST_UPD created_on,
    c.LAST_UPD_BY lastupdatedby,
    c.ROW_ID "KEY",
    coalesce('<b>Name:</b> '+c.NAME,'<null>')
```

```
            +coalesce(',<br><b>Catalog/Category:</b> '+c4.NAME,'<null>') SOLUTION,
        c.DESC_TEXT description,
        c.NAME title,
        c.NAME name,
        c.FILE_REV_NUM +'' + c3.LIT_ID + ''+ c3.CTLG_CAT_ID + ''+ c4.ROW_ID + ''
            + c4.NAME summary,
        c.LIT_CD "type",
        c.BU_ID visibilityid,
        c4.NAME sblvisibilityinfo,
        'Sales Tool' sblbctype
FROM
    dbo.S_LIT c
    INNER JOIN      dbo.S_CTLG_CAT_LIT c3 ON c3.LIT_ID     = c.ROW_ID
    INNER JOIN      dbo.S_CTLG_CAT      c4 ON c4.ROW_ID     = c3.CTLG_CAT_ID
    LEFT OUTER JOIN dbo.S_CONTACT       c5 ON c5.PAR_ROW_ID = c.LAST_UPD_BY
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT 'callcenter_
enu/start.swe?SWECmd=GotoView&SWEView=All+Sales+Tools+across+Organizations&SWERF=1
&SWEHo=HostName&SWEBU=1&SWEApplet0=Sales+Tool+List+Applet&SWERowId0='+c.ROW_ID
URL, 'US' LANG, c.LAST_UPD LASTMODIFIEDDATE, 'text/html' CONTENTTYPE, c.LAST_UPD
created_on, c.LAST_UPD_BY lastupdatedby, c.ROW_ID "KEY", coalesce('<b>Name:</b>
'+c.NAME,'<null>')+coalesce(',<br><b>Catalog/Category:</b> '+c4.NAME,'<null>')
SOLUTION, c.DESC_TEXT description, c.NAME title, c.NAME name, c.FILE_REV_NUM +'' +
c3.LIT_ID + ''+ c3.CTLG_CAT_ID + ''+ c4.ROW_ID + '' + c4.NAME summary, c.LIT_CD
"type", c.BU_ID visibilityid, c4.NAME sblvisibilityinfo, 'Sales Tool' sblbctype
FROM dbo.S_LIT c INNER JOIN dbo.S_CTLG_CAT_LIT c3 ON c3.LIT_ID=c.ROW_ID INNER JOIN
dbo.S_CTLG_CAT c4 ON c4.ROW_ID=c3.CTLG_CAT_ID LEFT OUTER JOIN dbo.S_CONTACT c5 ON
c5.PAR_ROW_ID=c.LAST_UPD_BY
```

Solution

```
SELECT
    'callcenter_enu/start.swe?SWECmd=GotoView
        &SWEView=All+Solution+List+View
        &SWERF=1
        &SWEHo=HostName
        &SWEBU=1
        &SWEApplet0=Solution+List+Applet
        &SWERowId0='+c.ROW_ID URL,
    'US' LANG,
    c.LAST_UPD LASTMODIFIEDDATE,
    'text/html' CONTENTTYPE,
    c.ROW_ID "KEY",
    coalesce('<b>Name:</b> '+c.NAME,'<null>')
        +coalesce(',<br><b>Catalog/Category: </b>'+t.NAME,'<null>')
        +coalesce(',<br><b>Question: </b>'+ cast(c.FAQ_QUES_TEXT
            as nvarchar(4000)),'<null>')
        +coalesce(',<br><b>Resolution: </b>'+ cast(c.RESOLUTION_TEXT
            as nvarchar(4000)),'<null>') SOLUTION,
    c.ROW_ID sblrowid,
    c.CREATED created_on,
    c.CREATED_BY createdby,
    c.NAME title,
    c.FAQ_QUES_TEXT description,
    c.RESOLUTION_TEXT summary,
    c.TYPE_CD "type",
    c.STATUS_CD status,
    usr.LOGIN owner,
```

```
      usr.LOGIN alias,
      t.NAME location,
      'Solution' sblbctype
FROM
   dbo.S_RESITEM c
   INNER JOIN dbo.S_USER          usr ON c.CREATED_BY  = usr.PAR_ROW_ID
   INNER JOIN dbo.S_CTLGCT_RESITM cct ON c.ROW_ID      = cct.RES_ITEM_ID
   INNER JOIN dbo.S_CTLG_CAT      t   ON t.ROW_ID      = cct.CTLG_CAT_ID
   INNER JOIN dbo.S_CONTACT       c2  ON c2.PAR_ROW_ID = c.LAST_UPD_BY
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT 'callcenter_
enu/start.swe?SWECmd=GotoView&SWEView=All+Solution+List+View&SWERF=1&SWEHo=HostNam
e&SWEBU=1&SWEApplet0=Solution+List+Applet&SWERowId0='+c.ROW_ID URL, 'US' LANG,
c.LAST_UPD LASTMODIFIEDDATE, 'text/html' CONTENTTYPE, c.ROW_ID "KEY",
coalesce('<b>Name:</b> '+c.NAME,'<null>')+coalesce(',<br><b>Catalog/Category:
</b>'+t.NAME,'<null>') + coalesce(',<br><b>Question: </b>'+ cast(c.FAQ_QUES_TEXT
as nvarchar(4000)),'<null>')+ coalesce(',<br><b>Resolution: </b>'+
cast(c.RESOLUTION_TEXT as nvarchar(4000)),'<null>') SOLUTION, c.ROW_ID sblrowid,
c.CREATED created_on, c.CREATED_BY createdby, c.NAME title, c.FAQ_QUES_TEXT
description, c.RESOLUTION_TEXT summary, c.TYPE_CD "type", c.STATUS_CD status,
usr.LOGIN owner, usr.LOGIN alias, t.NAME location, 'Solution' sblbctype FROM
dbo.S_RESITEM c INNER JOIN dbo.S_USER usr ON c.CREATED_BY = usr.PAR_ROW_ID INNER
JOIN dbo.S_CTLGCT_RESITM cct ON c.ROW_ID = cct.RES_ITEM_ID INNER JOIN dbo.S_CTLG_
CAT t ON t.ROW_ID = cct.CTLG_CAT_ID INNER JOIN dbo.S_CONTACT c2 ON c2.PAR_ROW_
ID=c.LAST_UPD_BY
```

Service Request

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=All+Service+Request+across+Organizations
      &SWERF=1
      &SWEHo=HostName
      &SWEBU=1
      &SWEApplet0=Service+Request+List+Applet
      &SWERowId0='+c.ROW_ID URL,
   'US' LANG,
   c.LAST_UPD LASTMODIFIEDDATE,
   'text/html' CONTENTTYPE,
   c.ROW_ID "KEY",
   coalesce('<b>SR Number:</b> '+c.SR_NUM,'<null>')
      +coalesce(',<br><b>Summary:</b> '+c.SR_TITLE,'<null>')
      +coalesce(',<br><b>Status:</b> '+c.SR_STAT_ID,'<null>')
      +coalesce(',<br><b>Area:</b> '+c.SR_AREA,'<null>')
      +coalesce(',<br><b>Subarea:</b> '+c.SR_SUB_AREA,'<null>')
      +coalesce(',<br><b>Resolution:</b> '+c.RESOLUTION_CD,'<null>') SOLUTION,
   c.DESC_TEXT description,
   c.BU_ID visibilityid,
   c.ROW_ID sblrowid,
   c.CREATED created_on,
   c.CREATED_BY createdby,
   c.SR_TITLE summary,
   a.NAME orgName,
   c.SR_AREA code01,
   a.OU_NUM csn,
   contact.FST_NAME firstName,
   contact.LAST_NAME lastName,
   c.SR_NUM title,
```

```
      c.SR_STAT_ID status,
      c.SR_SUB_AREA code02,
      usr.LOGIN owner,
      'Service Request' sblbctype
FROM
      dbo.S_ORG_EXT a
      INNER JOIN      dbo.S_SRV_REQ c        ON a.PAR_ROW_ID       = c.CST_OU_ID
      LEFT OUTER JOIN dbo.S_CONTACT contact  ON contact.PAR_ROW_ID = c.CST_CON_ID
      LEFT OUTER JOIN dbo.S_USER    usr      ON usr.PAR_ROW_ID     = c.OWNER_EMP_ID
      LEFT OUTER JOIN dbo.S_CONTACT c2       ON c2.PAR_ROW_ID      = c.LAST_UPD_BY
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT 'callcenter_
enu/start.swe?SWECmd=GotoView&SWEView=All+Service+Request+across+Organizations&SWE
RF=1&SWEHo=HostName&SWEBU=1&SWEApplet0=Service+Request+List+Applet&SWERowId0='+c.R
OW_ID URL, 'US' LANG, c.LAST_UPD LASTMODIFIEDDATE, 'text/html' CONTENTTYPE, c.ROW_
ID "KEY", coalesce('<b>SR Number:</b> '+c.SR_
NUM,'<null>')+coalesce(',<br><b>Summary:</b> '+c.SR_
TITLE,'<null>')+coalesce(',<br><b>Status:</b> '+c.SR_STAT_
ID,'<null>')+coalesce(',<br><b>Area:</b> '+c.SR_
AREA,'<null>')+coalesce(',<br><b>Subarea:</b> '+c.SR_SUB_
AREA,'<null>')+coalesce(',<br><b>Resolution:</b> '+c.RESOLUTION_CD,'<null>')
SOLUTION, c.DESC_TEXT description, c.BU_ID visibilityid, c.ROW_ID sblrowid,
c.CREATED created_on, c.CREATED_BY createdby, c.SR_TITLE summary, a.NAME orgName,
c.SR_AREA code01, a.OU_NUM csn, contact.FST_NAME firstName, contact.LAST_NAME
lastName, c.SR_NUM title, c.SR_STAT_ID status, c.SR_SUB_AREA code02, usr.LOGIN
owner, 'Service Request' sblbctype FROM dbo.S_ORG_EXT a INNER JOIN dbo.S_SRV_REQ c
ON a.PAR_ROW_ID= c.CST_OU_ID LEFT OUTER JOIN dbo.S_CONTACT contact ON contact.PAR_
ROW_ID =c.CST_CON_ID LEFT OUTER JOIN dbo.S_USER usr ON usr.PAR_ROW_ID = c.OWNER_
EMP_ID LEFT OUTER JOIN dbo.S_CONTACT c2 ON c2.PAR_ROW_ID=c.LAST_UPD_BY
```

Contacts

```
SELECT
      'callcenter_enu/start.swe?SWECmd=GotoView
          &SWEView=All+Contacts+across+Organizations
          &SWERF=1
          &SWEHo=HostName
          &SWEBU=1
          &SWEApplet0=Contact+List+Applet
          &SWERowId0='+c.PAR_ROW_ID URL,
      'US' LANG,
      c.LAST_UPD LASTMODIFIEDDATE,
      'text/html' CONTENTTYPE,
      c.PAR_ROW_ID "KEY",
      '<b>Name: </b>'
          +coalesce(c.LAST_NAME,'<null>')+' '
          +coalesce(c.FST_NAME,'<null>')
          +coalesce(',<br><b>Phone No.:</b> '+c.WORK_PH_NUM,'<null>')
          +coalesce(',<br><b>E-Mail ID:</b> '+ c.EMAIL_ADDR,'<null>') SOLUTION,
      t.PERS_AGENDA agenda,
      c.PAR_ROW_ID sblrowid,
      c.CREATED created_on,
      c.CREATED_BY createdby,
      a.NAME+'#'+c.JOB_TITLE PATH,
      c.LAST_NAME+' '+c.FST_NAME title,
      c.LAST_NAME lastName,
      c.FST_NAME firstName,
      c.EMP_ID owner,
```

```
      c.EMAIL_ADDR emailID,
      c.WORK_PH_NUM phoneNumber02,
      'Contacts' sblbctype,
      t.ACCOMPLISH summary,
      addr.ZIPCODE zipcode,
      addr.COUNTRY country,
      party.NAME name,
      addr.ADDR street,
      c.BU_ID visibilityid
FROM
   dbo.S_PARTY party
   INNER JOIN      dbo.S_CONTACT c     ON party.ROW_ID  = c.PAR_ROW_ID
   INNER JOIN      dbo.S_POSTN_CON T3  ON c.PR_POSTN_ID = T3.POSTN_ID
     AND c.ROW_ID = T3.CON_ID
   INNER JOIN      dbo.S_PARTY T4      ON T3.POSTN_ID   = T4.ROW_ID
   LEFT OUTER JOIN dbo.S_ORG_EXT a     ON a.PAR_ROW_ID  = c.PR_DEPT_OU_ID
   LEFT OUTER JOIN dbo.S_ADDR_ORG addr ON addr.ROW_ID   = c.PR_PER_ADDR_ID
   LEFT OUTER JOIN dbo.S_CONTACT_T t   ON c.ROW_ID      = t.PAR_ROW_ID
   LEFT OUTER join dbo.S_CONTACT c2    ON c2.ROW_ID     = c.LAST_UPD_BY
WHERE
   (c.PRIV_FLG = 'N')
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT 'callcenter_
enu/start.swe?SWECmd=GotoView&SWEView=All+Contacts+across+Organizations&SWERF=1&SW
EHo=HostName&SWEBU=1&SWEApplet0=Contact+List+Applet&SWERowId0='+c.PAR_ROW_ID URL,
'US' LANG, c.LAST_UPD LASTMODIFIEDDATE, 'text/html' CONTENTTYPE, c.PAR_ROW_ID
"KEY", '<b>Name: </b>'+coalesce(c.LAST_NAME,'<null>')+' '+coalesce(c.FST_
NAME,'<null>')+coalesce(',<br><b>Phone No.:</b> '+c.WORK_PH_
NUM,'<null>')+coalesce(',<br><b>E-Mail ID:</b> '+ c.EMAIL_ADDR,'<null>') SOLUTION,
t.PERS_AGENDA agenda, c.PAR_ROW_ID sblrowid, c.CREATED created_on, c.CREATED_BY
createdby, a.NAME+'#'+c.JOB_TITLE PATH, c.LAST_NAME+' '+c.FST_NAME title, c.LAST_
NAME lastName, c.FST_NAME firstName, c.EMP_ID owner, c.EMAIL_ADDR emailID, c.WORK_
PH_NUM phoneNumber02, 'Contacts' sblbctype, t.ACCOMPLISH summary, addr.ZIPCODE
zipcode, addr.COUNTRY country, party.NAME name, addr.ADDR street, c.BU_ID
visibilityid FROM dbo.S_PARTY party INNER JOIN dbo.S_CONTACT c ON party.ROW_ID =
c.PAR_ROW_ID INNER JOIN dbo.S_POSTN_CON T3 ON c.PR_POSTN_ID = T3.POSTN_ID AND
c.ROW_ID = T3.CON_ID INNER JOIN dbo.S_PARTY T4 ON T3.POSTN_ID = T4.ROW_ID LEFT
OUTER JOIN dbo.S_ORG_EXT a ON a.PAR_ROW_ID = c.PR_DEPT_OU_ID LEFT OUTER JOIN
dbo.S_ADDR_ORG addr ON addr.ROW_ID = c.PR_PER_ADDR_ID LEFT OUTER JOIN dbo.S_
CONTACT_T t ON c.ROW_ID=t.PAR_ROW_ID LEFT OUTER join dbo.S_CONTACT c2 ON c2.ROW_
ID=c.LAST_UPD_BY WHERE (c.PRIV_FLG = 'N')
```

Activity

For the queries shown in this section, before starting the crawl, you must replace *HostName* with the name or IP address of the Siebel host computer.

The following query can be used with Oracle Database:

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=All+Activity+List+View
      &SWERF=1
      &SWEHo=HostName
      &SWEBU=1
      &SWEApplet0=Activity+Form+Applet
      &SWERowId0='+T1.ROW_ID URL,
   'US' LANG,
```

```
              T1.ROW_ID "KEY",
              T1.LAST_UPD LASTMODIFIEDDATE,
              T1.NAME title,
              coalesce('Activity Name:'+T1.NAME,'<null>')
                  +coalesce(', Activity Type:'+T1.TODO_CD,'<null>')
                  +coalesce(', Creation Date:'+convert(varchar,T1.CREATED,103),'<null>')
                  +coalesce(', Activity Status:'+T1.EVT_STAT_CD,'<null>')
                  +coalesce(', ActivityPriority:'+T1.EVT_PRIORITY_CD,'<null>')
                  +coalesce(', Activity Owner: '+T1.OWNER_LOGIN,'<null>') CONTENT,
              T1.NAME ActivityName,
              T1.CREATED CreatedOn,
              T1.CREATED_BY CreatedBy,
              T1.LAST_UPD_BY LastUpdatedBy,
              T1.TODO_CD ActivityType,
              T1.EVT_STAT_CD ActivityStatus,
              T1.EVT_PRIORITY_CD ActivityPriority,
              T1.TARGET_OU_ID Organization,
              T1.OWNER_LOGIN visibilityid,
              T1.OWNER_LOGIN ActivityOwner,
              T1.ROW_ID SBLROWID,
              T1.TODO_PLAN_START_DT StartDate,
              T1.TODO_PLAN_END_DT EndDate,
              T1.TODO_DUE_DT DueDate
FROM
       dbo.S_EVT_ACT T1
       INNER JOIN      dbo.S_ORG_EXT T2  ON T2.PAR_ROW_ID = T1.TARGET_OU_ID
       LEFT OUTER JOIN dbo.S_USER    usr ON usr.LOGIN     = T1.OWNER_LOGIN
WHERE
       ((T1.APPT_REPT_REPL_CD IS NULL)
       AND ((T1.TEMPLATE_FLG != 'Y')
       OR (T1.TEMPLATE_FLG = NULL)))
```

The following is the same query for use with Oracle Database, but formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT 'callcenter_
enu/start.swe?SWECmd=GotoView&SWEView=All+Activity+List+View&SWERF=1&SWEHo=HostNam
e&SWEBU=1&SWEApplet0=Activity+Form+Applet&SWERowId0='+T1.ROW_ID URL,'US'
LANG,T1.ROW_ID "KEY", T1.LAST_UPD LASTMODIFIEDDATE, T1.NAME title ,
coalesce('Activity Name:'+T1.NAME,'<null>')
+coalesce(', Activity Type:'+T1.TODO_CD,'<null>')
+coalesce(', Creation Date:'+convert(varchar,T1.CREATED,103),'<null>')
+coalesce(', Activity Status:'+T1.EVT_STAT_CD,'<null>')
+coalesce(', ActivityPriority:'+T1.EVT_PRIORITY_CD,'<null>')
+coalesce(', Activity Owner: '+T1.OWNER_LOGIN,'<null>') CONTENT,T1.NAME
ActivityName,T1.CREATED CreatedOn, T1.CREATED_BY CreatedBy,T1.LAST_UPD_BY
LastUpdatedBy, T1.TODO_CD ActivityType,  T1.EVT_STAT_CD ActivityStatus,T1.EVT_
PRIORITY_CD ActivityPriority, T1.TARGET_OU_ID Organization,T1.OWNER_LOGIN
visibilityid, T1.OWNER_LOGIN ActivityOwner,T1.ROW_ID SBLROWID,T1.TODO_PLAN_START_
DT StartDate, T1.TODO_PLAN_END_DT EndDate, T1.TODO_DUE_DT DueDate FROM dbo.S_EVT_
ACT T1 INNER JOIN dbo.S_ORG_EXT T2 ON T2.PAR_ROW_ID = T1.TARGET_OU_ID LEFT OUTER
JOIN dbo.S_USER usr ON usr.LOGIN = T1.OWNER_LOGIN WHERE ((T1.APPT_REPT_REPL_CD IS
NULL) AND ((T1.TEMPLATE_FLG != 'Y') OR (T1.TEMPLATE_FLG = NULL)))
```

The following query can be used with SQL Server:

```
SELECT
       'callcenter_enu/start.swe?SWECmd=GotoView
          &SWEView=All+Activity+List+View
          &SWERF=1
          &SWEHo=HostName
```

```
        &SWEBU=1
        &SWEApplet0=Activity+Form+Applet
        &SWERowId0='+T1.ROW_ID URL,
    'US' LANG,
    T1.ROW_ID "KEY",
    T1.LAST_UPD LASTMODIFIEDDATE,
    T1.NAME title,
    coalesce('Activity Name:'+T1.NAME,'<null>')
        +coalesce(', Activity Type:'+T1.TODO_CD,'<null>')
        +coalesce(', Creation Date:'+convert(varchar,T1.CREATED,103),'<null>')
        +coalesce(', Activity Status:'+T1.EVT_STAT_CD,'<null>')
        +coalesce(', ActivityPriority:'+T1.EVT_PRIORITY_CD,'<null>')
        +coalesce(', Activity Owner: '+T1.OWNER_LOGIN,'<null>') CONTENT,
    T1.NAME ActivityName,
    T1.CREATED CreatedOn,
    T1.CREATED_BY CreatedBy,
    T1.LAST_UPD_BY LastUpdatedBy,
    T1.TODO_CD ActivityType,
    T1.EVT_STAT_CD ActivityStatus,
    T1.EVT_PRIORITY_CD ActivityPriority,
    T1.TARGET_OU_ID Organization,
    T1.OWNER_LOGIN visibilityid,
    T1.OWNER_LOGIN ActivityOwner,
    T1.ROW_ID SBLROWID,
    T1.TODO_PLAN_START_DT StartDate,
    T1.TODO_PLAN_END_DT EndDate,
    T1.TODO_DUE_DT DueDate
FROM
    dbo.S_EVT_ACT T1
    INNER JOIN dbo.S_ORG_EXT   T2  ON T2.PAR_ROW_ID = T1.TARGET_OU_ID
    LEFT OUTER JOIN dbo.S_USER usr ON usr.LOGIN     = T1.OWNER_LOGIN
WHERE
    ((T1.APPT_REPT_REPL_CD IS NULL)
    AND ((T1.TEMPLATE_FLG != 'Y')
    OR (T1.TEMPLATE_FLG = NULL)))
```

The following is the same query for use with SQL Server, but formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT 'callcenter_
enu/start.swe?SWECmd=GotoView&SWEView=All+Activity+List+View&SWERF=1&SWEHo=HostNam
e&SWEBU=1&SWEApplet0=Activity+Form+Applet&SWERowId0='+T1.ROW_ID URL,'US'
LANG,T1.ROW_ID "KEY", T1.LAST_UPD LASTMODIFIEDDATE, T1.NAME title,
coalesce('Activity Name:'+T1.NAME,'<null>')
+coalesce(', Activity Type:'+T1.TODO_CD,'<null>')
+coalesce(', Creation Date:'+convert(varchar,T1.CREATED,103),'<null>')
+coalesce(', Activity Status:'+T1.EVT_STAT_CD,'<null>')
+coalesce(', ActivityPriority:'+T1.EVT_PRIORITY_CD,'<null>')
+coalesce(', Activity Owner: '+T1.OWNER_LOGIN,'<null>') CONTENT, T1.NAME
ActivityName,T1.CREATED CreatedOn, T1.CREATED_BY CreatedBy,T1.LAST_UPD_BY
LastUpdatedBy, T1.TODO_CD ActivityType,  T1.EVT_STAT_CD ActivityStatus,T1.EVT_
PRIORITY_CD ActivityPriority, T1.TARGET_OU_ID Organization,T1.OWNER_LOGIN
visibilityid, T1.OWNER_LOGIN ActivityOwner,T1.ROW_ID SBLROWID, T1.TODO_PLAN_START_
DT StartDate, T1.TODO_PLAN_END_DT EndDate,  T1.TODO_DUE_DT DueDate  FROM dbo.S_
EVT_ACT T1 INNER JOIN dbo.S_ORG_EXT T2 ON T2.PAR_ROW_ID = T1.TARGET_OU_ID LEFT
OUTER JOIN dbo.S_USER usr ON usr.LOGIN = T1.OWNER_LOGIN WHERE ((T1.APPT_REPT_REPL_
CD IS NULL) AND ((T1.TEMPLATE_FLG != 'Y') OR (T1.TEMPLATE_FLG = NULL)))
```

Activity Attachment

For the queries shown in this section, before starting the crawl, substitute the following with the appropriate values:

- *Siebel_Host*: Siebel host name or IP address.

- *Folder_Path*: Absolute path to the folder containing attachment files on the Siebel host.

- *Attachment_Decompressor*: Choose the appropriate value:

  - When crawling from Linux, UNIX, or from different Windows domain, use RMI:

    ```
    rmi://Siebel_Host[:RMI_Port]/Decompressor
    ```

    For example:

    ```
    rmi://ses5-pc.example.com:2155/Decompressor
    ```

    RMI_Port is the RMI port number. Specify it only when the RMI registry in the Siebel host is running on a port other than the default port 1099.

    The RMI Decompression Server must be running.

  - When crawling from Windows in the same domain, use Windows Native Access:

    ```
    \\Siebel_Host\sseunzip_path
    ```

    For example:

    ```
    \\ses5-pc.example.com\BIN\sseunzip.exe
    ```

    sseunzip_path is the path to the shared sseunzip.exe utility.

Use this query with Oracle Database:

```
SELECT
    'callcenter_enu/start.swe?SWECmd=GotoView
        &SWEView=Activity+Attachment+View
        &SWERF=1
        &SWEHo=Host_name
        &SWEBU=1
        &SWEApplet0=Activity+Form+Applet
        &SWERowId0='¦¦T2.ROW_ID¦¦'
        &SWEApplet1=Activity+Attachment+Applet
        &SWERowId1='¦¦T1.ROW_ID URL,
    'US' LANG,
    T1.ROW_ID KEY,
    GREATEST(T1.LAST_UPD, T2.LAST_UPD) LASTMODIFIEDDATE,
    DECODE(T1.FILE_EXT, 'ppt', 'application/vnd.ms-powerpoint',
        'doc', 'application/msword', 'html', 'text/html',
        'txt', 'text/plain', 'pdf', 'application/pdf','xls',
        'application/vnd.ms-excel') CONTENTTYPE,
    Attachment_Decompressor
    Folder_Path\S_ACTIVITY_ATT_'+T1.ROW_ID+'_'
        +FILE_REV_NUM+'.SAF' ATTACHMENT_LINK,
    T1.FILE_NAME CONTENT,
    T1.FILE_NAME TITLE,
    T1.CREATED CreatedOn,
    T1.CREATED_BY CreatedBy,
    T1.LAST_UPD_BY LastUpdatedBy,
    T1.FILE_SRC_TYPE FileType,
```

```
    T1.FILE_DATE FileDate,
    replace(T1.FILE_NAME,' ','_') FileName,
    T1.FILE_SIZE FileSize,
    T1.FILE_EXT FileExtension,
    T1.COMMENTS AttachmentComment,
    'Activity Attachment' sblbctype,
    T2.OWNER_LOGIN ActivityOwner,
    T2.OWNER_LOGIN visibilityid,
    T2.TARGET_OU_ID Organization,
    T2.ROW_ID SBLROWID
FROM
    dbo.S_ACTIVITY_ATT T1
    INNER JOIN      dbo.S_EVT_ACT T2  ON T1.PAR_ROW_ID = T2.ROW_ID
    LEFT OUTER JOIN dbo.S_USER    usr ON usr.LOGIN      = T2.OWNER_LOGIN
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT 'callcenter_
enu/start.swe?SWECmd=GotoView&SWEView=Activity+Attachment+View&SWERF=1&SWEHo=Host_
name&SWEBU=1&SWEApplet0=Activity+Form+Applet&SWERowId0='¦¦T2.ROW_
ID¦¦'&SWEApplet1=Activity+Attachment+Applet&SWERowId1='¦¦T1.ROW_ID URL,'US'
LANG,T1.ROW_ID KEY,GREATEST(T1.LAST_UPD, T2.LAST_UPD) LASTMODIFIEDDATE,
decode(T1.FILE_EXT, 'ppt', 'application/vnd.ms-powerpoint', 'doc',
'application/msword', 'html', 'text/html', 'txt', 'text/plain', 'pdf',
'application/pdf','xls', 'application/vnd.ms-excel') CONTENTTYPE, 'Attachment_
Decompressor Folder_Path\S_ACTIVITY_ATT_'+T1.ROW_ID+'_'+FILE_REV_NUM+'.SAF'
ATTACHMENT_LINK,T1.FILE_NAME CONTENT, T1.FILE_NAME TITLE, T1.CREATED
CreatedOn,T1.CREATED_BY CreatedBy,T1.LAST_UPD_BY LastUpdatedBy,T1.FILE_SRC_TYPE
FileType,T1.FILE_DATE FileDate,replace(T1.FILE_NAME,' ','_') FileName,T1.FILE_SIZE
FileSize, T1.FILE_EXT FileExtension, T1.COMMENTS AttachmentComment,'Activity
Attachment' sblbctype,T2.OWNER_LOGIN ActivityOwner,T2.OWNER_LOGIN
visibilityid,T2.TARGET_OU_ID Organization, T2.ROW_ID SBLROWID FROM dbo.S_ACTIVITY_
ATT T1 INNER JOIN dbo.S_EVT_ACT T2 ON T1.PAR_ROW_ID=T2.ROW_ID LEFT OUTER JOIN
dbo.S_USER usr ON usr.LOGIN = T2.OWNER_LOGIN
```

Use this query with SQL Server:

```
SELECT
    'callcenter_enu/start.swe?SWECmd=GotoView
        &SWEView=Activity+Attachment+View
        &SWERF=1
        &SWEHo=Host_name
        &SWEBU=1
        &SWEApplet0=Activity+Form+Applet
        &SWERowId0='+T2.ROW_ID+'
        &SWEApplet1=Activity+Attachment+Applet
        &SWERowId1='+T1.ROW_ID URL,
    'US' LANG,
    T1.ROW_ID "KEY",
    CASE
        WHEN (DATEDIFF(second, T1.LAST_UPD, T2.LAST_UPD)> 0) THEN T2.LAST_UPD
        ELSE T1.LAST_UPD
    END
LASTMODIFIEDDATE,
    Attachment_Decompressor Folder_Path
        \S_ACTIVITY_ATT_'+T1.ROW_ID+'_'+FILE_REV_NUM+'.SAF' ATTACHMENT_LINK,
    T1.FILE_NAME CONTENT,
    T1.FILE_NAME TITLE,
    T1.CREATED CreatedOn,
    T1.CREATED_BY CreatedBy,
```

```
                    T1.LAST_UPD_BY LastUpdatedBy,
                    T1.FILE_SRC_TYPE FileType,
                    T1.FILE_DATE FileDate,
                    REPLACE(T1.FILE_NAME,' ','_') FileName,
                    T1.FILE_SIZE FileSize,
                    T1.FILE_EXT FileExtension,
                    T1.COMMENTS AttachmentComment,
                    'Activity Attachment' sblbctype,
                    T2.OWNER_LOGIN ActivityOwner,
                    T2.OWNER_LOGIN visibilityid,
                    T2.TARGET_OU_ID Organization,
                    T2.ROW_ID SBLROWID,
                    CASE T1.FILE_EXT
                        WHEN 'ppt'  THEN 'application/vnd.ms-powerpoint'
                        WHEN 'doc'  THEN 'application/msword'
                        WHEN 'html' THEN 'text/html' when 'txt' then 'text/plain'
                        WHEN 'pdf'  THEN 'application/pdf'
                        WHEN 'xls'  THEN 'application/vnd.ms-excel'
                    END
                    CONTENTTYPE
FROM
    dbo.S_ACTIVITY_ATT T1
    INNER JOIN      dbo.S_EVT_ACT T2  ON T1.PAR_ROW_ID = T2.ROW_ID
    LEFT OUTER JOIN dbo.S_USER     usr ON usr.LOGIN     = T2.OWNER_LOGIN
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT 'callcenter_
enu/start.swe?SWECmd=GotoView&SWEView=Activity+Attachment+View&SWERF=1&SWEHo=Host_
name&SWEBU=1&SWEApplet0=Activity+Form+Applet&SWERowId0='+T2.ROW_
ID+'&SWEApplet1=Activity+Attachment+Applet&SWERowId1='+T1.ROW_ID URL,'US'
LANG,T1.ROW_ID "KEY",case when (datediff(second, T1.LAST_UPD, T2.LAST_UPD)> 0)
then T2.LAST_UPD else T1.LAST_UPD end LASTMODIFIEDDATE,'Attachment_Decompressor
Folder_Path\S_ACTIVITY_ATT_'+T1.ROW_ID+'_'+FILE_REV_NUM+'.SAF' ATTACHMENT_LINK,
T1.FILE_NAME CONTENT, T1.FILE_NAME TITLE, T1.CREATED CreatedOn,T1.CREATED_BY
CreatedBy, T1.LAST_UPD_BY LastUpdatedBy,T1.FILE_SRC_TYPE FileType,T1.FILE_DATE
FileDate,replace(T1.FILE_NAME,' ','_') FileName,T1.FILE_SIZE FileSize, T1.FILE_EXT
FileExtension, T1.COMMENTS AttachmentComment,'Activity Attachment'
sblbctype,T2.OWNER_LOGIN ActivityOwner,T2.OWNER_LOGIN visibilityid,T2.TARGET_OU_ID
Organization, T2.ROW_ID SBLROWID,  case T1.FILE_EXT when 'ppt' then
'application/vnd.ms-powerpoint' when 'doc' then 'application/msword' when 'html'
then 'text/html' when 'txt' then 'text/plain' when 'pdf' then 'application/pdf'
when 'xls' then 'application/vnd.ms-excel' end CONTENTTYPE FROM dbo.S_ACTIVITY_ATT
T1 INNER JOIN dbo.S_EVT_ACT T2 ON T1.PAR_ROW_ID=T2.ROW_ID LEFT OUTER JOIN dbo.S_
USER usr ON usr.LOGIN = T2.OWNER_LOGIN
```

Setting Up Siebel 8 Sources

The Siebel 8 connector uses the Oracle SES XML connector framework, where searching is based on Siebel data available as XML feeds.

See Also:

- "Overview of XML Connector Framework" on page 4-11

- *Siebel Search Administration Guide*:

  http://docs.oracle.com/cd/B40099_
  02/books/Search/SearchTOC.html

- Siebel documentation on Oracle Technology Network (OTN) for
  information about supported Siebel modules:

  http://www.oracle.com/technetwork/indexes/documentation/i
  ndex.html

To activate an identity plug-in for Siebel 8 sources:

1. On the Global Settings page, select **Identity Management Setup**.

2. Select **Siebel 8** and click **Activate** to display the Activate Identity Plug-in page.

3. Enter values for the parameters as described in Table 9–11. Obtain these values
   from the Siebel administrator.

4. Click **Finish**.

Table 9–11   Siebel 8 Identity Management Parameters

Parameter	Value
Siebel 8 Authentication Web Service Endpoint	HTTP endpoint of the Siebel Web service that provides the authentication service
Siebel 8 Validation Web Service Endpoint	HTTP endpoint of the Siebel Web service that provides the user validation service
User ID	Administrator ID for accessing the user validation service
Password	User password.

To create a Siebel 8 source:

1. Activate an identity plug-in as described in the previous procedure.

2. On the home page, select the **Sources** secondary tab.

3. Select **Siebel 8** from the Source Types list, and click **Create**.

4. Enter the source parameters as described in Table 9–12.

5. Click **Next**.

6. Click **Get Parameters** to obtain a list of parameters for the authorization manager
   plug-in.

7. Enter the values for the authorization manager plug-in parameters as described in
   Table 9–13.

8. Click **Create**.

Table 9–12   Siebel 8 Source Parameters

Parameter	Value
Configuration URL	FILE protocol address of the XML configuration file providing details about the source, such as the data feed type, location, security attributes, and so on.
	Obtain this file from Siebel administrator and save it on the same computer as Oracle SES. Enter the configuration URL in the form:
	`file://localhost/`*`config_path`* where *`config_path`* is the absolute path to the configuration file. For example: `file://localhost/private/oracle/config.xml/`
Authentication Type	Standard Java authentication type used by the application serving the control and data feed. Leave this parameter blank as the feeds are accessed over file or FTP protocols.
User ID	User ID to login to the FTP server and access the data feeds. The access details of the data feed are specified in the configuration file. The user id can be obtained from Siebel administrator.
Password	Password for **User ID**.
Realm	The realm of the application serving the feeds. Leave this parameter blank since the feeds are accessed over file or FTP.
Oracle SSO Login URL	Set this parameter value when the authentication type is ORASSO. Oracle SES redirects the crawler to this SSO login URL for authentication before crawling the Siebel 8 source.
	The format for this URL for Oracle 10g SSO secured Siebel 8 source is:
	■ For basic authentication type:
	`https://`*`server`*`:`*`port`*`/pls/orasso/orasso.wwsso_app_ admin.ls_login`
	■ For form based authentication type:
	`https://`*`server`*`:`*`port`*`/mysso/signon.jsp`
	The format for this URL for Oracle 11g OAM secured Siebel 8 source is:
	`http://`*`server`*`:`*`port`*`/oam/server/obrareq.cgi?encquery`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.
Oracle SSO Action URL	Set this parameter value when the authentication type is ORASSO. This is the URL that is displayed after successfully authenticating the SSO user before crawling the Siebel 8 source.
	The format for this URL for Oracle 10g SSO secured Siebel 8 source is:
	`https://`*`server`*`:`*`port`*`/sso/auth`
	The format for this URL for Oracle 11g OAM secured Siebel 8 source is:
	`http://`*`server`*`:`*`port`*`/oam/server/auth_cred_submit`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.
Scratch Directory	A directory on the same computer as Oracle SES, where the status logs are created temporarily.
Maximum number of connection attempts	Maximum number of attempts to connect to the target server to access the data feed.

Table 9–12   (Cont.) Siebel 8 Source Parameters

Parameter	Value
Delete Linked Documents	Set it to true if the documents crawled from the links in the feeds must be deleted. Set it to false otherwise. Default is true.
Number of data feeds to be pre-fetched	Number of data feeds to pre-fetch. Its value can be any number greater than or equal to 0. If its value is set to 0, then no data feed is pre-fetched. The value specified must be such that there is sufficient memory to cache the pre-fetched feeds. The default value is 0.
Enable Resume Crawl	Set it to true to enable resuming crawl, when the crawl fails or stops. The default value is true.
Stop Crawl On Content Fetch Error	Set it to true to stop crawl on content fetch error, else set it to false. The default value is false.

Table 9–13   Siebel 8 Authorization Parameters

Parameter	Value
Siebel 8 authorization Web service endpoint	Webs service endpoint of the Siebel Web service that provides the authorization service.
User ID	Administrator ID for accessing the authorization service.
Password	Password for **User ID**.

Setting Up PeopleSoft Sources

Using Oracle SES, you can search for documents within PeopleSoft application. This is done by establishing a connection between Oracle SES and PeopleSoft using the PeopleSoft connector. To connect to and retrieve documents from PeopleSoft, you must set up an Oracle SES PeopleSoft identity management system using an identity plug-in, and an authorization management system using an authorization plug-in.

The identity plug-in enables Oracle SES to identify the set of users that can access the PeopleSoft application. The authorization plug-in enables Oracle SES to determine the access rights that each user has for accessing different documents and data within PeoPleSoft application. Usually, all users may not have access to the entire data and document set within the application. Instead, each user may have access to a limited set of documents and data.

Setting Up an Identity Management System

An identity management system enables Oracle SES to identify the set of users that can access the PeopleSoft application. An identity management system is implemented using an identity plug-in in Oracle SES.

To activate an identity plug-in for PeopleSoft sources:

1. On the Global Settings page, click **Identity Management Setup** to open the Identity Management Setup page.

2. From the list of available sources, select **PeopleSoft**, and click **Activate**.

   This opens the Activate Identity Plug-in page.

3. Enter values as described in Table 9–14. Obtain the values from the PeopleSoft application administrator.

4. Click **Finish**.

Table 9–14   Identity Management Parameters for PeopleSoft

Parameter	Description
HTTP end point for authentication	The HTTP endpoint to which user authentication/validation requests are sent.
User ID	Administration user ID to be used in the HTTP request for user authentication. This user ID is used to validate the authentication request in the PeopleSoft repository. Obtain this ID from the PeopleSoft application administrator.
Password	Administration password.

Defining a PeopleSoft Source

A PeopleSoft application source can be defined from the Source page. After you define the source, you can search for documents within the application.

To create a PeopleSoft source:

1. On the home page, click the **Sources** subtab.

    This opens the Sources page.

2. From Source Type list, select **PeopleSoft** and click **Create**.

    This opens the Create Source page, which guides you through a multi-step procedure to enter source and authorization parameters.

3. On the Create Source page, enter the source parameter values listed in Table 9–15.

4. Click **Next** and specify values for the authorization parameters listed in Table 9–16.

5. Click **Create & Customize** to create the source.

Table 9–15   PeopleSoft Source Parameters

Parameter	Description
Configuration URL	URL of the XML configuration file providing details of the source, such as the data feed type, location, security attributes, and so on. Obtain this file from the PeopleSoft application administrator.
Authentication Type	Standard Java authentication type used by the application. Enter `BASIC` for basic authentication, `FORM` for form-based authentication, `ORASSO` for Oracle SSO, `NATIVE` for proprietary XML over HTTP authentication.
User ID	User ID to access the data feeds. The access details of the data feed are specified in the configuration file. The user id can be obtained from PeopleSoft administrator.
Password	User password.
Realm	The realm of the application serving the feeds. Enter the relevant information when the feeds are accessed over HTTP and is mandatory when the authentication type is `BASIC`.

Table 9–15   (Cont.) PeopleSoft Source Parameters

Parameter	Description
Oracle SSO Login URL	Set this parameter value when the authentication type is ORASSO. Oracle SES redirects the crawler to this SSO login URL for authentication before crawling the PeopleSoft source.
	The format for this URL for Oracle 10g SSO secured PeopleSoft source is:
	■ For basic authentication type:
	`https://`*server*`:`*port*`/pls/orasso/orasso.wwsso_app_` `admin.ls_login`
	■ For form based authentication type:
	`https://`*server*`:`*port*`/mysso/signon.jsp`
	The format for this URL for Oracle 11g OAM secured PeopleSoft source is:
	`http://`*server*`:`*port*`/oam/server/obrareq.cgi?encquery`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.
Oracle SSO Action URL	Set this parameter value when the authentication type is ORASSO. This is the URL that is displayed after successfully authenticating the SSO user before crawling the PeopleSoft source.
	The format for this URL for Oracle 10g SSO secured PeopleSoft source is:
	`https://`*server*`:`*port*`/sso/auth`
	The format for this URL for Oracle 11g OAM secured PeopleSoft source is:
	`http://`*server*`:`*port*`/oam/server/auth_cred_submit`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.
Scratch Directory	Local directory where status files can be temporarily written.
Maximum number of connection attempts	Maximum number of connection attempts to access data feed or upload status feed.
Delete Linked Documents	Set it to `true` if the documents crawled from the links in the feeds must be deleted. Set it to `false` otherwise. Default is `true`.
Number of data feeds to be pre-fetched	Number of data feeds to pre-fetch. Its value can be any number greater than or equal to 0. If its value is set to 0, then no data feed is pre-fetched. The value specified must be such that there is sufficient memory to cache the pre-fetched feeds. The default value is 0.
Enable Resume Crawl	Set it to `true` to enable resuming crawl, when the crawl fails or stops. Set it to `false` otherwise. Default is `true`.
Stop Crawl On Content Fetch Error	Set it to `true` to stop crawl on content fetch error, else set it to `false`. The default value is `false`.

Table 9–16   PeopleSoft Connector Authorization Parameters

Parameter	Description
HTTP endpoint for authorization	HTTP endpoint for PeopleSoft authorization. For example, `http://`*my.host.com:port*`/AppSearch/SecurityService`
User ID	Administration user ID for PeopleSoft authorization.

Table 9–16 (Cont.) PeopleSoft Connector Authorization Parameters

Parameter	Description
Password	Administration password.
Business component	Name of PeopleSoft Business Component. For example, `oracle.apps.fnd.fwk.search.NavigationSVO`
Display URL Prefix	HTTP host to prefix the access URL to form the display URL. For example, `http://my.host.com:7777/`. This value must form a valid URL when concatenated with the access URL element of an item in the data feed. Be careful to avoid having either two slashes or none when the values are combined. Thus, enter a trailing slash (/) if the access URLs do not begin with a slash, or omit the trailing slash from the prefix if the access URLs begin with a slash.
Security attribute values for anonymous user	Comma-delimited list of authorized values of security attributes for anonymous user. When this parameter is left blank, the authorization service is contacted to retrieve the values of security attributes accessible for anonymous users.
User Identity Format	Format of user identity string posted to PeopleSoft Authorization service. The default value is `FND`.

Setting Up Oracle Fusion Sources

Using Oracle SES, you can search for documents within Oracle Fusion Applications. This is done by establishing a connection between Oracle SES and Oracle Fusion using a Fusion connector. To connect to and retrieve documents from Oracle Fusion, you must set up an Oracle SES Fusion identity management system using an identity plug-in, and an authorization management system using an authorization plug-in.

The identity plug-in enables Oracle SES to identify the set of users that can access the Fusion application. The authorization plug-in enables Oracle SES to determine the access rights that each user has for accessing different documents and data within Oracle Fusion Applications. Usually, all users may not have access to the entire data and document set within the application. Instead, each user may have access to a limited set of documents and data.

Setting Up an Identity Management System

An identity management system enables Oracle SES to identify the set of users that can access the Fusion application. An identity management system is implemented using an identity plug-in in Oracle SES.

To activate an identity plug-in for Fusion sources:

1. On the Global Settings page, click **Identity Management Setup** to open the Identity Management Setup page.

2. From the list of available sources, select **Oracle Fusion**, and click **Activate**.

   This opens the Activate Identity Plug-in page.

3. Enter values as described in Table 9–17. Obtain the values from the Fusion application administrator.

4. Click **Finish**.

Table 9–17   Identity Management Parameters for Oracle Fusion

Parameter	Description
HTTP end point for authentication	The HTTP endpoint to which user authentication/validation requests are sent.
User ID	Administration user ID to be used in the HTTP request for user authentication. This user ID is used to validate the authentication request in the Fusion repository. Obtain this ID from the Fusion application administrator.
Password	Administration password.

Defining a Fusion Source

A Fusion application source can be defined from the Source page. After you define the source, you can search for documents within the application.

To create a Fusion source:

1. On the home page, click the **Sources** subtab.

   This opens the Sources page.

2. From Source Type list, select **Oracle Fusion** and click **Create**.

   This opens the Create Source page, which guides you through a multi-step procedure to enter source and authorization parameters.

3. On the Create Source page, enter the source parameter values listed in Table 9–18.

4. Click **Next** and specify values for the authorization parameters listed in Table 9–19.

5. Click **Create & Customize** to create the source.

Table 9–18   Oracle Fusion Source Parameters

Parameter	Description
Configuration URL	URL of the XML configuration file providing details of the source, such as the data feed type, location, security attributes, and so on. The URL is a HTTP URL accessible over HTTP. Obtain this file from the Fusion application administrator.
Authentication Type	Enter the value `NATIVE`.
User ID	User ID to access the data feeds. The access details of the data feed are specified in the configuration file. The user id can be obtained from Fusion administrator.
Password	User password.
Realm	The realm of the application serving the feeds. The parameter is usually left blank.

Table 9–18   (Cont.) Oracle Fusion Source Parameters

Parameter	Description
Oracle SSO Login URL	Set this parameter value when the authentication type is ORASSO. Oracle SES redirects the crawler to this SSO login URL for authentication before crawling the Oracle Fusion source.
	The format for this URL for Oracle 10g SSO secured Oracle Fusion source is:
	■ For basic authentication type:
	`https://server:port/pls/orasso/orasso.wwsso_app_ admin.ls_login`
	■ For form based authentication type:
	`https://server:port/mysso/signon.jsp`
	The format for this URL for Oracle 11g OAM secured Oracle Fusion source is:
	`http://server:port/oam/server/obrareq.cgi?encquery`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.
Oracle SSO Action URL	Set this parameter value when the authentication type is ORASSO. This is the URL that is displayed after successfully authenticating the SSO user before crawling the Oracle Fusion source.
	The format for this URL for Oracle 10g SSO secured Oracle Fusion source is:
	`https://server:port/sso/auth`
	The format for this URL for Oracle 11g OAM secured Oracle Fusion source is:
	`http://server:port/oam/server/auth_cred_submit`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.
Scratch Directory	Local directory where status files can be temporarily written.
Maximum number of connection attempts	Maximum number of connection attempts to access data feed or upload status feed.
Delete Linked Documents	Set it to `true` if the documents crawled from the links in the feeds must be deleted. Set it to `false` otherwise. Default is `true`.
Number of data feeds to be pre-fetched	Number of data feeds to pre-fetch. Its value can be any number greater than or equal to 0. If its value is set to 0, then no data feed is pre-fetched. The value specified must be such that there is sufficient memory to cache the pre-fetched feeds. The default value is 0.
Enable Resume Crawl	Set it to `true` to enable resuming crawl when the crawl fails or stops. The default value is `true`.
Stop Crawl On Content Fetch Error	Set it to `true` to stop crawl on content fetch error, else set it to `false`. The default value is `false`.

Table 9–19   Fusion Connector Authorization Parameters

Parameter	Description
HTTP endpoint for authorization	HTTP endpoint for Oracle Fusion authorization. For example, `http://my.host.com:port/AppSearch/SecurityService`
User ID	Administration user ID for Oracle Fusion authorization.

Table 9–19   (Cont.) Fusion Connector Authorization Parameters

Parameter	Description
Password	Administration password.
Business component	Name of Oracle Fusion Business Component. For example, `oracle.apps.fnd.fwk.search.NavigationSVO`
Display URL Prefix	HTTP host to prefix the access URL to form the display URL. For example, `http://my.host.com:7777/`.
	This value must form a valid URL when concatenated with the access URL element of an item in the data feed. Be careful to avoid having either two slashes or none when the values are combined. Thus, enter a trailing slash (/) if the access URLs do not begin with a slash, or omit the trailing slash from the prefix if the access URLs begin with a slash.
Security attribute values for anonymous user	Comma-delimited list of authorized values of security attributes for anonymous user. When this parameter is left blank, the authorization service is contacted to retrieve the values of security attributes accessible for anonymous users.
User Identity Format	Format of user identity string posted to Oracle Fusion Authorization service. The default value is `FND`.

Setting Up Oracle WebCenter Sources

Use the WebCenter connector to connect to and search for documents within Oracle WebCenter 11*g*.

To set up WebCenter connector, you must define source parameters for the connector and set up an identity management system using an identity plug-in.

Setting Up an Identity Management System

An identity management system enables Oracle SES to identify the set of users that can access the WebCenter application. An identity management system is implemented using an identity plug-in in Oracle SES. A WebCenter source can use identity management systems, such as, Oracle Internet Directory (OID), Active Directory, or Oracle Unified Directory (OUD).

The following example describes the steps to activate Sun Java System Directory Server identity plug-in for using Oracle Unified Directory (OUD) identity management system for WebCenter sources.

To activate Sun Java System Directory Server identity plug-in for using OUD for WebCenter sources:

1. On the Global Settings page, click **Identity Management Setup** to open the Identity Management Setup page.

2. Select **Sun Java System Directory Server**, and click **Activate**.

    This opens the Activate Identity Plug-in page.

3. Enter the values as described in Table 9–20 for using the OUD identity management system for WebCenter sources. Obtain these values from the WebCenter application administrator.

4. Click **Finish**.

Table 9–20   Identity Management Parameters for using Oracle Unified Directory (OUD) for WebCenter

Parameter	Description
Directory URL	Directory Server URL. For example, `ldap://OUD_Directory_Server:4389`
Directory account name	Directory Server account to connect to. For example, `cn=Directory Manager`
Directory account password	Directory Server account password.
Directory subscriber	Directory Server subscriber (LDAP based). For example, `dc=us,dc=org,dc=com` for the domain name `us.org.com`
Directory security protocol	Set it to `none`.
Directory user cache	Set it to `false`.

Defining a WebCenter Source

A WebCenter source can be defined from the Source page. After you define the source, you can search for documents within the application.

To create a WebCenter source:

1. On the home page, click the **Sources** subtab.

   This opens the Sources page.

2. From Source Type list, select **Oracle WebCenter** and click **Create**.

   This opens the Create Source page, which guides you through a multi-step procedure to enter source and authorization parameters.

3. On the Create Source page, enter the source parameter values listed in Table 9–21.

4. Click **Next** and specify values for the authorization parameters listed in Table 9–22.

5. Click **Create & Customize** to create the source.

Table 9–21   Oracle WebCenter Source Parameters

Parameter	Description
Configuration URL	URL of the XML configuration file providing details of the source, such as the data feed type, location, security attributes, and so on. The URL is a HTTP URL accessible over HTTP. Obtain this file from the WebCenter application administrator.
Authentication Type	Enter the value `NATIVE`.
User ID	User ID to access the data feeds. The access details of the data feed are specified in the configuration file. The user id can be obtained from Fusion administrator.
Password	User password.
Realm	The realm of the application serving the feeds. The parameter is usually left blank.

Table 9–21   (Cont.) Oracle WebCenter Source Parameters

Parameter	Description
Oracle SSO Login URL	Set this parameter value when the authentication type is ORASSO. Oracle SES redirects the crawler to this SSO login URL for authentication before crawling the WebCenter source.
	The format for this URL for Oracle 10g SSO secured WebCenter source is:
	■ For basic authentication type:
	`https://`*server*`:`*port*`/pls/orasso/orasso.wwsso_app_admin.ls_login`
	■ For form based authentication type:
	`https://`*server*`:`*port*`/mysso/signon.jsp`
	The format for this URL for Oracle 11g OAM secured WebCenter source is:
	`http://`*server*`:`*port*`/oam/server/obrareq.cgi?encquery`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.
Oracle SSO Action URL	Set this parameter value when the authentication type is ORASSO. This is the URL that is displayed after successfully authenticating the SSO user before crawling the WebCenter source.
	The format for this URL for Oracle 10g SSO secured WebCenter source is:
	`https://`*server*`:`*port*`/sso/auth`
	The format for this URL for Oracle 11g OAM secured WebCenter source is:
	`http://`*server*`:`*port*`/oam/server/auth_cred_submit`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.
Scratch Directory	Local directory where status files can be temporarily written.
Maximum number of connection attempts	Maximum number of connection attempts to access data feed or upload status feed.
Delete Linked Documents	Set it to `true` if the documents crawled from the links in the feeds must be deleted. Set it to `false` otherwise. Default is `true`.
Number of data feeds to be pre-fetched	Number of data feeds to pre-fetch. Its value can be any number greater than or equal to 0. If its value is set to 0, then no data feed is pre-fetched. The value specified must be such that there is sufficient memory to cache the pre-fetched feeds. The default value is 0.
Enable Resume Crawl	Set it to `true` to resume crawl when the crawl fails or stops; else set it to `false`. The default value is `true`.
Stop Crawl On Content Fetch Error	Set it to `true` to stop crawl on content fetch error, else set it to `false`. The default value is `false`.

Table 9–22   WebCenter Connector Authorization Parameters

Parameter	Description
HTTP Endpoint for Authorization	URL servicing the lookup of authorization information.

Table 9–22   (Cont.) WebCenter Connector Authorization Parameters

Parameter	Description
Display URL Prefix	HTTP host information to prefix the relative access URL to form the complete display URL. For example, `http://my.host.com:7777/idc` This value must form a valid URL when concatenated with the access URL element of an item in the data feed. Be careful to avoid having either two slashes or none when the values are combined. Thus, enter a trailing slash (/) if the access URLs do not begin with a slash, or omit the trailing slash from the prefix if the access URLs begin with a slash.
Administrator User	Administrator user ID to authenticate to the authorization URL.
Administrator Password	Administrator user password to authenticate to the authorization URL.
Authorization User ID Format	Format of the active identity plug-in user ID that is used by the WebCenter authorization endpoint. For example, this can be the username, email ID, or nickname.
Display Crawled Version	Controls access to the crawled documents: ■ `true`: Search results point to the crawled version of the document. ■ `false`: Search results point to the content information page.
Authorization User ID Format	Format of the user ID used by the WebCenter authorization API, such as `username`, `email`, `nickname`, `user_name`.
Use Cached User and Role Information to Authorize Results	Controls user authorization: ■ `true`: Uses the cached user query filter. This setting removes the query time dependency on WebCenter. ■ `false`: Queries WebCenter for authorization.
User Role Data Source to Cache the Filter	The name of the WebCenter Users source that has crawled the user's SecurityGroup and Account information.
Realm	Realm of the application serving the authorization information.

Oracle SSO Login URL	Set this parameter value when the authentication type is ORASSO. Oracle SES redirects the crawler to this SSO login URL for authentication before crawling the WebCenter source.
	
	The format for this URL for Oracle 10g SSO secured WebCenter source is:
	
	■ For basic authentication type:
	
	`https://server:port/pls/orasso/orasso.wwsso_app_`
	`admin.ls_login`
	
	■ For form based authentication type:
	
	`https://server:port/mysso/signon.jsp`
	
	The format for this URL for Oracle 11g OAM secured WebCenter source is:
	
	`http://server:port/oam/server/obrareq.cgi?encquery`
	
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.

Table 9–22   (Cont.) WebCenter Connector Authorization Parameters

Parameter	Description
Oracle SSO Action URL	Set this parameter value when the authentication type is ORASSO. This is the URL that is displayed after successfully authenticating the SSO user before crawling the WebCenter source.
	The format for this URL for Oracle 10g SSO secured WebCenter source is:
	`https://server:port/sso/auth`
	The format for this URL for Oracle 11g OAM secured WebCenter source is:
	`http://server:port/oam/server/auth_cred_submit`
	The value for *port* can be found from the "Listen Port" field in the OAM Managed Server Console.

Part III

Advanced Topics

This part provides information for experienced administrators. It contains the following chapters:

- Chapter 10, "Customizing the Search User Interface"
- Chapter 11, "Security in Oracle Secure Enterprise Search"
- Chapter 12, "Administering Oracle SES Instances"
- Chapter 13, "Oracle Secure Enterprise Search APIs"

10

Customizing the Search User Interface

You can customize the Oracle SES UI by including new template skins. This chapter consists of the following topic:

- Overview of the Search GUI
- Using Skin Bundles to Customize the Search GUI

Overview of the Search GUI

You can customize the Oracle SES query UI by modifying certain template files that define the display content in the UI. These template files can then be used by the application logic to render the Oracle SES query UI. You can also create style sheets (CSS files) that define how the pages are visually styled and then provide a link to the CSS files from the template files.

To customize the UI using the template files and style sheets, you must know how to use the FreeMarker Template language and CSS style sheets. See the FreeMarker Web site for information about FreeMarker template files.

The changes that can be made to the UI include:

- Changing the logo
- Changing the look and feel including the colors and fonts
- Modifying the page header and footer including static text and links to other pages
- Supporting multiple skins for the UI. These skins can be selected using a URL parameter.
- Supporting RSS, Atom, and CSV formats on the result list.
- Using common header and footer templates that can be included by all pages.
- Modifying certain behavior of Oracle SES JavaScript components, such as result clustering.

The following aspects of the Query UI cannot be customized using template files and style sheets:

- The search results list, which is handled by the Search Results List Advanced Configuration Style Sheets.
- The Oracle SES query engine full-text search pipeline. The search query on the text index is performed first. Then, the results data is used to render the search results using customized template files.

■ The internal rendering logic for the various visual components on the search results pages. These components are provided as FreeMarker macros in the template library.

User Interface Components

Oracle SES provides a set of user interface components in a template library called `seslib.ftl`. You can import this library into the template files. To use the library in a template file, insert the following line at the beginning of the template file:

```
<#import "/lib/oracle.com/seslib.ftl" as ses>
```

The user interface components are implemented using macros in FreeMarker.

These components help in separating the code that defines the internal operations of the Oracle SES query engine from the code that defines the display layout of the UI. You can thus alter the text and modify the display layout without editing any of the code files that control the query engine (.jsp files).

The UI components are listed in Figure 10–1.

Figure 10–1   The UI Components of the Search Results Page
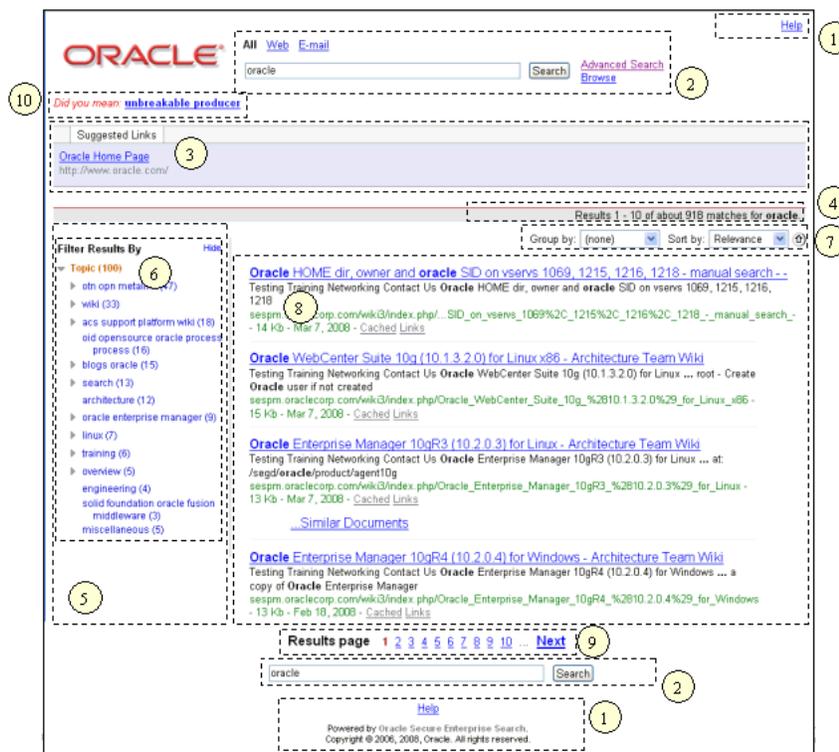
The user interface components include:

1. Header and footer (`inc_header.ftl`, `inc_footer.ftl`)

2. Search query box

3. Suggested links and suggested content display panel

4. Context bar subcomponents including bread crumbs and hit statistics

5. Expandable sidebar

6. Result cluster trees

7. Sorting and grouping controls

8. Result list

9. Pagination links

10. Alternate keywords display

Editable Template Files

Oracle SES contains template files that are written in the FreeMarker template language (.ftl). To customize the UI, you can modify these template files. Table 10–1 lists the template files and the corresponding query application pages where they are included. For more information on editing the template files, see *Oracle Secure Enterprise Search Administration API Guide*.

Table 10–1   List of Template Files

UI Pages	Description	Template Files to Customize the UI Pages
Main search pages	These pages display the query box, search results, and any search errors and offer the most customization possibilities.	The following template files are used to customize the main search pages: ■ query.ftl: The initial page displayed before a user enters a search query. Can be customized to display a splash screen. ■ results.ftl: Displays the search results page. ■ noresults.ftl: This page is displayed when a search query returns no results. You can customize it to provide useful query suggestions to the user. ■ error.ftl: Displays an error message for invalid search queries, such as invalid syntax. ■ inc_header.ftl, inc_footer.ftl: These template files are used to customize the header and footer information of the main search pages
Supplemental pages	These pages include the ■ Links page: display_links.jsp ■ E-mail result browser: mail.jsp and pmail.jsp ■ Customize pages: customize.jsp ■ Submit URL page: submit_url.jsp ■ Login page: login.jsp	The header, footer, and CSS files for the supplemental pages can be customized by modifying the following template files:inc_supplemental_header.ftl, inc_supplemental_footer.ftl, and inc_supplemental_includes.ftl

Table 10–1   (Cont.) List of Template Files

UI Pages	Description	Template Files to Customize the UI Pages
Feeds and CSV	These output pages display the search results as RSS feeds or in a comma-delimited format.	These pages can be customized using the following template files: ■ `atom.ftl`: Provides an ATOM feed of the search results, which are returned as text or XML from the `feed` page. This is done by specifying the option `feed?output=atom`. The default template renders the data in the ATOM Feed 0.3 format. ■ `rss.ftl`: Provides an RSS feed of the search results. The results are returned in the text or XML format from the `feed` page. The default template renders the data in the RSS 2.0 format. ■ `csv.ftl`: This provides the search results in a comma-delimited format. The results are returned in the text format from the `csv` page.

Template File for Additional Error Handling

In addition to the template files listed in Table 10–1, there is a template file called `unchecked_error.ftl`, which is used for error handling. This template file is used when there is a severe error and the application throws an unexpected exception. In such cases, the normal UI consisting of the source group tabs, query box, browse popup, and cluster sidebar is not displayed. Instead, the display page consists of the error message, and the header and footer. This template file also includes the supplemental template files `inc_supplemental_includes.ftl`, `inc_supplemental_header.ftl`, and `inc_supplemental_footer.ftl`, to render the header and footer.

Data Model for the Template Files

The data model of a template file defines the dynamic data that is used to render the view.

Data Model for the Main Search Pages

The data model for the files `query.ftl`, `results.ftl`, `noresults.ftl`, and `error.ftl` contains the variables listed in Table 10–2:

Table 10–2   Data Models for the Main Search Pages

Variable	Description
authn	This hash contains information related to secure login and authentication.
authn.isLoggedIn	Boolean value that indicates whether the current user is logged in.

Table 10–2   (Cont.) Data Models for the Main Search Pages

Variable	Description
authn.loggedInUser.displayName	If isloggedIn is set to `true`, then this parameter holds the user name. If isLoggedIn is set to `false`, then this parameter does not hold any value.
authn.isAuthenticationActive	Boolean value that indicates whether secure search is enabled.
params	This hash contains the URL query parameters for the request. It is an `HttpRequestParameterHashModel`, as provided by the FreeMarker API.
htmlLang	The language for the `<html>` tag.
skin	The current skin name.
sourceGroups[]	This sequence is a list of hashes, each representing one source group active on the instance.
sourceGroups[n].id	The source group ID.
sourceGroups[n].name	The source group untranslated name.
sourceGroups[n].displayName	The translated source group name.
pageUrls	This hash contains variables that contain the URL of pages in the query application.
pageUrls.search	the URL for `/search`.
pageUrls.help	The URL for online help.
pageUrls.login	The login URL, if the user is not logged in.
pageUrls.logout	The logout URL, if the user is logged in.
pageUrls.customize	The Personalize URL.
pageUrls.submitUrl	The URL for the Submit URL page.
showSidebar	Boolean value that indicates whether the result clustering sidebar should be shown.
errorMessage	If an error occurred, this variable holds the error message.
req	This hash represents the search query request.
req.displayQuery	The search query to be displayed, usually for the page title and the hit statistics.
req.queryboxQuery	The portion of the search query entered in the input box, excluding attribute filters.
res	This hash represents the search query response.
res.queryId	The query ID, used for click statistics.
res.startNum	The index of the first result document returned.
res.endNum	The index of the last result document returned.
res.docsReturned	The number of documents returned with this request.
res.estimatedHitCount	The estimated total of all documents that match the query.
res.suggestedLinks[]	This sequence is a list of hashes, each representing a suggested links.
res.suggestedLinks[n].title	The title to display for the suggested link.
res.suggestedLinks[n].url	The URL of the suggested link.

Table 10–2   (Cont.) Data Models for the Main Search Pages

Variable	Description
res.resultElements[]	This sequence is a list of hashes, each representing a result document. The format of each result document is not exposed.
perf	This includes basic performance metrics.
perf.reqTime	The total request time from the start of the JSP processing to immediately before template rendering.
perf.searchTime	The time spent in the Java midtier query and backend layers for the text query.
perf.templateLoadTime	The time spent loading the FreeMarker template from the ftl file.

Data Model for the Supplemental Pages

The data model for the supplemental pages is as listed in Table 10–3:

Table 10–3   Data Models for Supplemental Pages

Variables	Description
authn	This hash contains information related to secure login and authentication.
authn.isLoggedIn	Boolean value that indicates whether the current user is logged in.
authn.loggedInUser.displayName	If `isloggedIn` is set to `true`, then this parameter holds the user name. If `isLoggedIn` is set to `false`, then this parameter does not hold any value.
authn.isAuthenticationActive	Boolean value that indicates whether secure search is enabled.
htmlLang	The language for the `<html>` tag
skin	The current skin name.
pageUrls	This hash contains variables that contain the URL of pages in the query app.
pageUrls.search	The URL for `/search`.
pageUrls.help	The URL for online help.
pageUrls.login	The login URL, if the user is not logged in.
pageUrls.logout	The logout URL, if the user is logged in.
pageUrls.customize	The Personalize URL.
pageUrls.submitUrl	The URL for the Submit URL page.

Data Model for Output Format Pages

The data model for the files `atom.ftl`, `rss.ftl`, and `csv.ftl` contains the following variables, as listed in Table 10–4:

Table 10–4   Data Model for Output Format Pages

Variables	Description
authn	This hash contains information related to secure login and authentication.

Table 10–4   (Cont.) Data Model for Output Format Pages

Variables	Description
authn.isLoggedIn	Boolean value that indicates whether the current user is logged in.
authn.loggedInUser.displayName	If `isloggedIn` is set to `true`, then this parameter holds the user name. If `isLoggedIn` is set to `false`, then this parameter does not hold any value.
authn.isAuthenticationActive	Boolean value that indicates whether secure search is enabled.
isError	Boolean value that indicates whether an error occurred. This would be equivalent to a situation in which the `error.ftl` template was invoked.
params	This hash contains the URL query parameters for the request. It is an `HttpRequestParameterHashModel`, as provided by the FreeMarker API.
currentDate	The current date.
requestURL	The request URL, as returned by `HttpServletRequest.getRequestURL()`.
requestURLParams	The request query string, as returned by `HttpServletRequest.getQueryString()`.
req	This hash represents the search query request.
req.displayQuery	The search query to be displayed, usually for the page title and the hit statistics.
req.queryboxQuery	The portion of the search query entered in the input box, excluding attribute filters.
res	This hash represents the search query response.
res.startNum	The index of the first result document returned.
res.endNum	The index of the last result document returned.
res.docsReturned	The number of documents returned with this request.
res.estimatedHitCount	The estimated total of all documents that match the query.
res.resultElements[]	This sequence is a list of hashes, each representing a result document. The format of each result document is not exposed.

Handling Errors and Exceptions

Errors can creep in while creating the template files. Errors can also occur due to incorrect query syntax from the user end.

Errors in Template Files

Simple syntax errors in the template files are logged in a file called `AdminServer_yyyy_MM_dd_hh_mm.log` located at `wls_domain_home/ses_domain_name/servers/AdminServer/logs`.

Severe exceptions are handled by the file `searcherr.jsp` which contains the template file `unchecked_error.ftl`. This page displays only the error message. It does not display the source group tabs, the query box, the clustering sidebar, or the login information.

End User Query Syntax Errors

Query syntax exceptions and other handled exceptions are rendered by the `error.ftl` file in `search.jsp`. For example, if a user enters a query that contains invalid syntax or is too broad, then the rendered page contains an error message and other information such as the source groups and the login information.

Configuring the JavaScript Components

The following JavaScript components can be configured to render different UI pages.

SearchView

This component manages the query form, page variables, source group selection, and the `search within` display options.

Constructor

```
ORACLE.ses.SearchView(obj config)
```

Key configuration properties

- **searchFormName**: Name of the query form.

Key events

- **clearBrowse**: When the link `clear` for the `Search within` option is clicked.

- **groupTabClick**: When a source group tab is clicked.

Result Cluster Tree

Configure the display options for the result cluster tree that is displayed in the search results page using the JavaScript component called `ClusterView`.

The constructor is

```
ORACLE.ses.ClusterView(str|HTMLElement el, obj config);
```

The key configuration properties are:

- **displayEmptyTrees**: Boolean that determines the display option of trees. Set as `true` to display grayed-out empty trees, or `false` to hide them.

- **autoExpandFirstLevel**: Boolean indicating whether the first level of each cluster tree should be expanded or not.

- **moreLimit**: Number of nodes in the first-level of each cluster tree before the `more` option is used.

- **highlightMode**: Whether to highlight the currently selected cluster node. It can take the values `OFF`, `NODE`, or `SUBTREE`. For example, the value can be `ORACLE.ses.ClusterView.HIGHLIGHT_MODE.SUBTREE`. Default value is `OFF`.

- **allowSelectionToggle**: A boolean that indicates whether clicking on a selected node toggles it off, or whether it remains selected. Default is `off/false`.

- **removeFilterId**: The ID of the link to `remove filter`.

The key events are:

- **beforeRenderClusterTree**: Fired before cluster tree is rendered.

- **renderClusterTree**: Fired after cluster tree is rendered.

- **beforeRenderClusterNode**: Fired before cluster node is rendered.

- **renderClusterNode**: Fired after cluster node is rendered.

- **beforeSelectNode**: Fired before cluster node is selected. Allows canceling selection.

- **selectNode**: Fired after cluster node is selected.

- **beforeDeselectNode**: Fired before cluster node is selected. Allows canceling selection.

- **deselectNode**: Fired after cluster node is deselected.

These events can be used for various functions such as changing the order of cluster nodes within a tree or relabeling cluster nodes. For example, changing the month label from 1 to January.

> **Note:** The `beforeDeselectNode` and `deselectNode` events are fired before any `selectNode` events. If deselection is canceled, the selection of the new node does not occur, because multiple selection is not allowed.

ResultsView and ResultsModel

This component asynchronously retrieves the top-N results, renders the results list, handles client-side caching of search results, pagination, and similar documents. Also handles grouping and sorting of the results list, besides asynchronous fetching of email threads.

Constructor

```
ORACLE.ses.ResultsView (obj SESConnectionFactory, obj config)
```

Key configuration properties

- **searchView**: The search view instance.

- **resultsListContainer**: The ID of the container element for the results list.

- **disableTopNResults**: Enables disabling the top-N results functionality while still loading the ResultsView component, for example, to support email threading.

- **paginatorContainers**: An array of elements or IDs of the containers for pagination links. If multiple containers are passed in, then they are updated together. This registers the paginators to be updated when the results change.

- **hitReportEl**: This registers the element to be updated when the results change.

- **loadingImg**: The ID of the loading icon image.

- **groupSortContextMsgEl**: An element representing the grouping or sorting context message. This registers the element to be updated when the results change. For example, if the `@topn_context_msg` macro is called directly to render the content message, then use the ID provided to the macro as the value for `groupSortContextMsgEl`.

- **groupingEl**: An element representing the grouping list. This registers the element to be updated when the results change. For example, if the `@grouping_control` macro is called directly to render the grouping list, then use the ID provided to the macro as the value for `groupingEl`.

- **sortingEl**: An element representing the sorting list and sort order toggle icon. This registers the element to be updated when the results change. For example, if the

@sorting_control macro is called directly to render the sorting list, then use the ID provided to the macro as the value for sortingEl.

Key events

- **beforeRenderGroupedResults**: Fired before grouped search results are rendered. Allows reordering of the groups.

- **modelChange**: Fired when any change occurs to the results model.

- **newResults**: Fired when new top-N results are retrieved.

InlineAdvancedView

This component displays the inline attribute filters table.

Constructor

```
ORACLE.ses.InlineAdvancedView(str|HTMLElement el, str attrNamesArrayID)
```

BrowseView and BrowseModel

Displays the **infosource** browse tree. You can control the display of the browse tree such that the entire infosource hierarchy is not loaded initially.

Constructor

```
ORACLE.ses.BrowseView(str|HTMLElement el, obj SESConnectionFactory, obj config)
```

Key configuration properties

- **prefetchBrowseGroups**: Boolean indicating whether to prefetch the top-level list of source groups, or load the list when the browse overlay is initially opened.

- **statusEl**: An element representing the browse tree status message, for example, Loading.

Key events

- **browseUpdate**: Fired when an update occurs to the browse model.

- **browseOff**: Fired when browse search within is switched off, for example, when all infosource nodes in the browse tree are deselected. As a result, the source group tabs are displayed.

- **allDocsClick**: Fired when a document count is clicked to view all the documents in that node.

SuggestedContentView and SuggestedContentModel

Displays suggested links and contents in a tabbed view.

Constructor

```
ORACLE.ses.SuggestedContentView(str|HTMLElement el, obj SESConnectionFactory, obj config)
```

Key configuration properties

- **encQuery**: The URL encoded query string.

SESConnectionFactory and SESConnection

Stores the base URL and any routing parameters to allow components to make an asynchronous requests to the Oracle SES server.

Constructor

```
ORACLE.ses.util.SESConnectionFactory(obj config)
```

Key configuration properties

- **baseurl**:The base URL of the Oracle SES application.

- **routingParams**: An array of routing parameters.

Key events

- `asyncRequest`: Supports the same options as
  `YAHOO.util.Connect.asyncRequest()`. Not however that the URI is relative to the
  base URL, and an additional options parameter allows passing in {serverAffinity:
  true} if routing parameters are needed for a particular asynchronous request.

AnchoredOverlay

The default usage of this component is to display the Browse Tree overlay next to the
`Browse` link.

Constructor

```
ORACLE.ses.AnchoredOverlay(str|HTMLElement el)
```

Key methods

- `setAnchor(str contextElId, str overlayCorner, str contextElCorner)`

- `render()`

Template File for Initializing JavaScript Components

In addition to the template files listed in Table 10–1, there is a template file called `inc_bootstrap_js.ftl` initializes the JavaScript components for the query application.
This includes initialization of the Search, Results, Browse, and Cluster components.

FreeMarker Macros

The template library file `seslib.ftl` contains the definitions of the following macros.
The output of these macros can be customized by changing the values of the input
arguments when the macros are called from the customized templates in a skin
bundle.

@navigation_links: Create a list of navigation links. Takes the following parameters:

- **separator**: The character or string to be displayed between each link.

- **preLinks, postLinks**: The list of links that must be displayed before and after the
  set of default links.

@logged_in_user: Displays the name of the logged in user.

@when_logged_in, **@when_not_logged_in**: You can specify content depending on if a
user is logged in or not. For example, for an anonymous user, you can set the
following display message:

```
<@when_not_logged_in>
 No secure results are shown.  Please log in to view secure results.
 </@when_not_logged_in>
```

@source_group_tabs, **@source_groups_dropdown**: To set the display of source groups
as tabs or as a drop-down list.

Note that either macro can be used in a page. However, you must not use both the macros together in a page.

The parameters are:

- **showDefaultAll:** Set this to `true` to display the default All group, or `false` to hide it.

- **suppressGroups**: The set of source groups that must not be displayed.

To suppress a multibyte source group, you must specify the source group name using its hexadecimal Unicode code (UCS code). Typically, UCS codes are of the format `\uxxxx`. For example `\u793e`. In the FreeMarker template file, you must replace `u` with `x`. For example, the UCS code `\u793e\u4f1a` must be specified as `\x793e\x4f1a`.

For example,

```
<@ses.source_group_tabs formName="${searchFormName}" suppressGroups
=["\x793e\x4f1a"]/>
```

> **Note:** There are many Web sites that provide multibyte character to UCS code conversion.

@source_groups_dropdown contains the following additional parameter:

- **submitOnChange**: Set this as true to submit the query form when a user selects a source group from the list.

@overlay: Renders the HTML for an overlay window. An overlay window can contain arbitrary content, for example, a browse tree or help content. This macro is intended to be used with the AnchoredOverlay JavaScript component. The parameters are:

- **id**: ID of the overlay window. This ID is passed in to the constructor of the `AnchoredOverlay` JavaScript component.

- **isVisible**: Set as true to display the overlay by default.

@feed_href, **@feed_icon**, **@feed_link_tag**: These are used to render ATOM or RSS feeds. `feed_icon` renders an icon for the feed. `feed_link_tag` renders an autodiscovery tag to the head section of the html.

The parameters for **@feed_href** are:

- **format**: Specifies whether it is an ATOM or RSS feed.

- **num**: The number of results to be displayed in the feed.

- **customAttrs**: List of names of attributes to include in the feed.

The **@feed_icon** parameters are:

- **img_src/width/height**: Provide the relative or absolute URL of the image, and its dimensions.

- **extra args**: Additional arguments may be used to add attributes to the anchor tag, for example to add a link target.

The **@feed_link_tag** parameters are:

- **format**: atom for ATOM feeds and rss for RSS feeds.

- **title**: The title attribute to use for the link tag.

For example, the following code displays a feed icon:

```
<@feed_icon img_src="images/icon/rss.gif"
```

```
target="_blank"><@feed_href customAttrs=["contributor", "category"]/>
</@feed_icon>
```

Similarly, to add a ATOM autodiscovery to a results page, add the following code to the template file `results.ftl`:

```
<html>
<head>
   <@feed_link_tag format="atom"><@feed_href/></@feed_link_tag>
...
</head>
...
</html>
```

@paginator: Renders pagination links for search results in the `results.ftl` template file. The parameters are:

- **formName**: The name of the query form.
- **id**: ID of the paginator container. It is used in the `paginatorContainers` argument when initializing the ResultsView JavaScript component.

@topn_status_icon: Displays a loading icon, error icon, or load icon depending on the state of the ResultsView JavaScript component.

@grouping_control: Displays the results grouping list.

@sorting_control: Displays the results sorting list along with the sort-order icon.

@topn_context_msg: Displays the on top NNN message.

The following parameters are common to the four macros:

- id: The ID of the control used in the ResultsView component.
- addClass: A CSS class to add to the control's `div` element. By default, the `@results_list_controls` macro adds the `controlBarItem` class to each individual control to manage styling and alignment.

@results_list_controls: This macro controls the four macros listed above. Use this to set the default values for all four macros.

> **Note:** ▪ If a customized arrangement of controls is used, then the corresponding parameter values that are used to register the controls with the `ResultsView` component should be changed or omitted. For example, if the grouping control is removed from a template, then the `groupingEl` parameter should not be passed in to the `ResultsView` component.
>
> ▪ Multibyte source group names must always be specified using the equivalent Unicode code (UCS code).

Using Skin Bundles to Customize the Search GUI

With Oracle SES 11g, the query UI can be customized using skin bundles, which are files grouped into a unit that create a customized skin for the search interface. A skin bundle may include JavaScript files, CSS and image files, and FreeMarker templates. See *Oracle Secure Enterprise Search Administration API Guide* for more information about skin bundles and using the Administration API to manage skin bundles.

In addition to using the Administration API, you can manage skin bundles using a WebDAV interface to the SES instance. The WebDAV interface enables you to track,

edit, and manage the skin bundles. You can add and update the skins including the template files and asset files through the WebDAV interface. See "About the WebDAV Interface" on page 10-15 for more information.

Customizing the Skin Bundles

In addition to versioned support bundles, Oracle SES also includes versioned skin bundles. Each skin is packaged as a skin bundle and contains the template files, and the asset files including CSS files and image files that define the skin. A skin bundle can be linked to a support bundle.

You can define the following properties for a skin bundle:

- **skinname**: Name of the skin.

- **linkedVersion**: The version of the Oracle SES UI support bundle that this skin is linked to, such as 11.2.2.2.0. If no version is specified when a skin bundle is created, the default value is the current major Oracle SES version. Note that the version number must consist of five numerals separated by dots (for example 11.2.2.2.0).

- **isDefault**: A flag to indicate whether a skin is the default skin. It can be set to `true` or `false`. If a user accesses the query application without specifying a skin name, then the skin bundle marked as default is used to render the UI. Only one skin bundle must be marked as the default skin. If no skin bundle is marked as default, then the default query UI is displayed, using the template files located in the current UI support bundle.

- **enabled**: A flag to indicate whether a skin is currently active in Oracle SES. It can be set to `true` or `false`. If set to `false`, then it is unavailable to end users. By default, the status for a skin is set to `false`.

For more information about these parameters, see "Configuring the Skin Bundle Properties" on page 10-16.

How Oracle SES Selects a Skin Bundle for Display

For any query request made by the user, Oracle SES selects an active skin bundle in the following order, terminating at the first match:

1. If a skin name is specified in the URL, then Oracle SES searches for the skin bundle. If the skin bundle is found and its state is set to `enabled`, then Oracle SES makes it the active skin bundle for this request.

2. If only one skin bundle is marked as default and it is set to `enabled`, then Oracle SES makes it the active skin bundle for this request.

3. If a skin bundle is not found based on the above criteria, then there are no active skin bundles. In such a scenario, Oracle SES displays the default UI support bundle.

> **Note:** Only one skin bundle must be marked as default using the `isDefault` parameter. However, using the WebDAV interface, it is possible to set multiple skin bundles as default. In such a scenario, Oracle SES cannot determine the default skin bundle. In such cases, Oracle SES uses the default support bundle to render the query result.

How Oracle SES Selects the Template Files

After Oracle SES determines the active skin bundle to be used for rendering the query request, the selection of the template files is done in the following order, terminating at the first match:

1. Oracle SES uses the template files available within the active skin bundle.

2. Else, it uses the template files available within the linked support bundle.

If template files are not available in either of the locations, then an error exception is thrown.

About the WebDAV Interface

The support bundles and the skin bundles are stored in the DB XML repository, and can be easily accessed over the WebDAV file access interface. To support WebDAV access, a listener port is enabled for the XML DB repository. This listener port enables the WebDAV HTTP port to communicate with the DB repository. The communication channel between the WebDAV interface and the repository is secured using an authentication mechanism. See "Securing the Oracle SES-WebDAV Communication Channel" on page 10-17 for more details.

Folder Structure for the Skin Bundles in the XML DB The support bundles and the skin bundles are stored in the DB repository. The top level folder is `skins` and is available at `/search/queryapp/skins/`.

The `skins` folder holds separate folders for each skin, that is `/search/queryapp/skins/`*`skinnames/`*.

Each skin folder contains subfolders called `templates` and `assets`, that is `/search/queryapp/skins/`*`skinname`*`/templates`

and

`/search/queryapp/skins/`*`skinname`*`/assets`

The `templates` folder contains the FreeMarker template files. The `assets` folder may hold JavaScript files, CSS style sheet files, and image files. Within the `templates` and `assets` folders, users can create their own subfolders. This is especially useful within the `assets` folder, where users can create separate subfolders to store images, style sheets, and JavaScript files.

For example, if you create a skin called `blue`, then the folder structure for the template file `results.ftl` is:

`/search/queryapp/skins/blue/templates/results.ftl`

and the path of a CSS file `blue.css` is:

`/search/queryapp/skins/blue/assets/blue.css`

or you can create a subfolder called `css` to store style sheets, in which case the path is:

`/search/queryapp/skins/blue/assets/css/blue.css`

Using FreeMarker Macros to Retrieve File Paths within Bundles FreeMarker macros enable you to retrieve the file paths of the various files that are available within skin bundles and support (product) bundles. This is useful when you need to reference any asset file from a template file. For example, you might need to reference an image file from a template file. In such cases, you are required to provide the relative path of the files, be it within a skin bundle or a support bundle.

The following macros enable you to output the relative URL paths of skin bundle asset files, support bundle JavaScript files, support bundle CSS files, and support bundle image files, respectively.

- **@skin_asset filename=**_file_name_: Retrieves the file location of any asset file located within the current skin bundle. It returns the URL of the file. This URL makes it accessible from a Web browser. For the `filename` parameter, you must specify the relative path of the asset file within the `/assets/`folder of the skin bundle.

- **@support_js filename=**_file_name_ **lib=**_lib_name_: Retrieves the location of any JavaScript file located within the current support bundle. It returns the URL of the file. This URL makes it accessible from a Web browser. For the `filename` parameter, you must specify the relative path of the JavaScript file, depending on the JavaScript library specified with the `lib` parameter. Valid values for the `lib` parameter are: `ses`, `yui`, and `bubbling`. The **@includes_page_head** macro located in the `seslib.ftl` template file contains relevant examples.

- **@support_css filename=**_file_name_ **lib=**_lib_name_: Retrieves the location of any CSS file located within the current support bundle. It returns the URL of the file. This URL makes it accessible from a Web browser. For the `filename` parameter, you must specify the relative path of the CSS file, depending on the library specified with the `lib` parameter. Valid values for the `lib` parameter are `ses` and `yui`. The **@includes_page_head** macro located in the `seslib.ftl` template file contains relevant examples.

- **@support_img filename=**_file_name_ **lib=**_lib_name_: Retrieves the location of any image file located within the current support bundle. It returns the URL of the file. This URL makes it accessible from a Web browser. For the filename parameter, you must specify the relative path of the image file, depending on the library specified with the `lib` parameter. Valid values for the `lib` parameter are `ses` and `yui`. The **@includes_page_head** macro located in the `seslib.ftl` template file contains relevant examples.

For example, consider a template file within the skin `blue`, deployed to the Oracle SES server `http://myhost.example.com:7777/search/query/`.

Suppose the template file contains the following template code to reference an asset file `arrow.gif`:

```
<img src="<@ses.skin_asset filename="images/arrow.gif"/>">
```

Then the html output includes the complete URL path for the image file `arrow.gif`. It is of the format:

```
<img
src="http://myhost.example.com:7777/search/query/assets/skins/blue/images/arrow.gi
f">
```

Configuring the Skin Bundle Properties   The parameters to configure a skin bundle are listed in an XML file called `skinconfig.xml`, which can be accessed and modified over the WebDAV interface. Any changes that you make to this file is immediately reflected in the Oracle SES query application.

This XML file must be placed directly within the folder for each skin. For example, for the skin `blue`, the `skinconfig.xml` file must be placed within `/search/queryapp/skins/blue/`.

The properties that can be set include the linked version, the `isDefault` flag, and the enabled/disabled state as described in "Customizing the Skin Bundles" on page 10-14.

The following example lists the contents of a `skinconfig.xml` file.

```
<skinconfig xmlns="http://xmlns.oracle.com/search/skinconfig.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://xmlns.oracle.com/search/skinconfig.xsd
http://xmlns.oracle.com/search/skinconfig.xsd">
  <isDefault>false</isDefault>
  <linkedVersion>11.2.2.2.0</linkedVersion>
  <enabled>true</enabled>
</skinConfig>
```

Securing the Oracle SES-WebDAV Communication Channel

Access to the Oracle SES application over the WebDAV interface is secured using an authentication mechanism, whereby only the `searchsys` user is allowed to access the skin bundle and support bundle folders.

By default, Oracle SES does not allow users to connect to listener ports from remote computers. This is controlled using the parameters `tcp.invited_nodes` and `tcp.validnode` in `oracle_home`/network/admin/sqlnet.ora.

To allow the `searchsys` user access to the WebDAV HTTP port from a remote computer, you must edit the following parameters in `sqlnet.ora` file as shown in the example:

```
tcp.invited_nodes=(testhost1,testhost2,127.0.0.1)
tcp.validnode_checking=yes
```

where `testhost1`, `testhost2`, and `127.0.0.1` are the IDs of remote computers that can access the listener.

> **Note:** Add the above parameters to the `sqlnet.ora` file, if they are not present.

Restart the Oracle SES middle tier and database after modifying `sqlnet.ora`.

Note that you can also provide the network address of the remote computer as the ID. If it is not desirable to open up the WebDAV port for remote access, then the Administration API can be used to manage skin bundles as an alternative to the WebDAV interface.

11

Security in Oracle Secure Enterprise Search

This chapter describes the architecture and configuration for Oracle SES security model. It contains the following topics:

- Overview of Oracle Secure Enterprise Search Security
- Enabling Secure Search
- Configuring Secure Search with OracleAS Single Sign-On
- Configuring Secure Search with Oracle Access Manager Single Sign-On
- SSL and HTTPS Support in Oracle Secure Enterprise Search
- Using Windows Native Authentication
- Changing the Master Encryption Key

Overview of Oracle Secure Enterprise Search Security

This section describes the Oracle SES security model. It contains the following topics:

- Oracle Secure Enterprise Search Security Model
- Changing the Administrator Password
- Authentication and Authorization
- Authentication Methods

Oracle Secure Enterprise Search Security Model

Oracle SES provides access to a variety of content repositories through a single gateway. Each external repository has its own security model that determines whether a particular user can access a particular document. You must carefully consider all aspects of security in Oracle SES to respect the security of documents coming from multiple data repositories.

Oracle SES uses the following security services in its security model:

- **Identity plug-ins** can obtain user and group information directly from any identity management system. An identity plug-in is Java code between Oracle SES and an identity management system, allowing Oracle SES to read user and group information.

- **Secure Socket Layers** (SSL) is the industry standard protocol for managing the security of message transmission on the Internet. This is used for securing RMI connections, HTTPS crawling, and secure JDBC.

Connecting to the Oracle SES server using SQL*Plus, except as documented in the guide, is not supported. Changing the Oracle SES server directly using SQL and modifying initialization parameter files is not supported. User management should only be done using the Oracle SES Administration GUI.

As an additional security measure, Oracle SES is configured to reject connection requests using SQL*Plus from remote hosts. The only protocols supported for connection to Oracle SES from remote hosts are HTTP, HTTPS, and AJP13.

Changing the Administrator Password

The user name for Oracle SES administrator is `searchsys`, who is also the WebLogic Server identity store user. The password for the `searchsys` user is specified during the installation. A password must consist of at least 8 characters, and it must contain at least one numeric and one alphabetic character. Note that the password length cannot exceed 30 characters. The password can only be changed using the WebLogic Server Administration Console.

To change the Oracle SES administrator password:

1. Log in to the WebLogic Server Administration Console.

2. Under the **Domain Structure** in the left panel, click **Security Realm**.

3. In the **Summary of Security Realms** section, click **myrealm**.

4. Click the **Users and Groups** tab.

5. Click **searchsys** from the list of users.

6. Click the **Passwords** tab.

7. Enter and confirm the new password.

8. Save the changes.

> **Note:** The `searchsys` user in the WebLogic Server identity store is different from the `searchsys` schema user of the Oracle SES database. Refer to the Oracle Database documentation for changing password for a Database schema user.

The following sections describe the steps for setting the Oracle SES users passwords in the WebLogic Server middle tier, if you change the passwords for the Oracle SES Database schema users `searchsys`, `search_ess`, and `search_mds`.

- Setting the SEARCHSYS Schema User Password in WebLogic Server Middle Tier

- Setting the SEARCH_ESS Schema User Password in WebLogic Server Middle Tier

- Setting the SEARCH_MDS Schema User Password in WebLogic Server Middle Tier

Setting the SEARCHSYS Schema User Password in WebLogic Server Middle Tier

If you change the password for the `searchsys` schema user of the Oracle SES database, then the following passwords related to the WebLogic Server configuration also need to be changed:

- Data source connection pool password

- Credential Storage Framework (CSF) password

> **Note:** If Oracle SES software is installed on an existing Database, then shut down the Oracle SES server and Oracle ESS server before executing the steps for changing the password related to the WebLogic Server configuration. Refer to "Starting and Stopping Oracle SES Instance" on page 2-6 for more information about starting and stopping WebLogic Server, Oracle ESS Server, and Oracle SES Server.

To change the data source connection pool password for SEARCHSYS user:

1. Log in to the WebLogic Server Administration Console.

2. Under the **Change Center** in the left panel, click **Lock & Edit**.

3. Change the password for **SearchAdminDS** data source:

   a. Under the **Domain Structure** in the left panel, select **search_domain > Services > Data Sources**. The Configuration tab for the settings for SearchAdminDS is displayed in the main panel.

   b. In the Name column of the Data Sources table, click **SearchAdminDS**. The General tab of the settings for SearchAdminDS is displayed.

   c. Select the **Connection Pool** tab.

   d. Enter and confirm the new password.

   e. Save the changes.

4. Change the password for **SearchQueryDS** data source:

   a. Under the **Domain Structure** in the left panel, select **search_domain > Services > Data Sources**. The Configuration tab for the settings for SearchQueryDS is displayed in the main panel.

   b. In the Name column of the Data Sources table, click **SearchQueryDS**. The General tab of the Settings for SearchQueryDS is displayed.

   c. Select the **Connection Pool** tab.

   d. Enter and confirm the new password.

   e. Save the changes.

5. Click **Activate Changes** under the **Change Center** in the left panel.

6. Restart the WebLogic Server.

To change the Credential Storage Framework (CSF) password for SEARCHSYS user:

1. Connect to the system where WebLogic Server middle tier is installed.

2. Go to *MW_HOME*/oracle_common/common/bin.

3. Start the WebLogic Server Administration Scripting Tool by executing the command `wlst.sh` on Linux and UNIX platform, and command `wlst.cmd` on Windows platform.

4. Run the `connect()` command at the `wls/offline>` prompt. You will be prompted to enter the following values:

   ▪ WebLogic Server user name

   ▪ WebLogic Server password

   ▪ WebLogic Server URL

5. Run the following command at the `wls:domain_name/serverConfig>` prompt:

```
updateCred(map="oracle.apps.security",key="FUSION_APPS_ECSF_SES_ADMIN-KEY",
user="searchsys", password="password")
```
where,

password is the SEARCHSYS user password, and the Oracle SES credentials for connecting to the Oracle Database are obtained from a CSF map named `oracle.apps.security` with a key named `FUSION_APPS_ECSF_SES_ADMIN-KEY`.

Setting the SEARCH_ESS Schema User Password in WebLogic Server Middle Tier

If you change the password for the `search_ess` schema user of Oracle SES database, then the data source connection pool password related to WebLogic Server configuration also needs to be changed.

> **Note:** If Oracle SES software is installed on an existing Database, then shut down the Oracle SES server and Oracle ESS server before executing the steps for changing the password related to the WebLogic Server configuration. Refer to "Starting and Stopping Oracle SES Instance" on page 2-6 for more information about starting and stopping WebLogic Server, Oracle ESS Server, and Oracle SES Server.

To change the data source connection pool password for SEARCH_ESS user:

1. Log in to the WebLogic Server Administration Console.

2. Under the **Change Center** in the left panel, click **Lock & Edit**.

3. Change the passwords for **EssDS, EssInternalDS**, and **EssXADS** data sources:

    a. Under the **Domain Structure** in the left panel, select **search_domain > Services > Data Sources**. The Configuration tabs for the settings for EssDS, EssInternalDS, and EssXADS is displayed in the main panel.

    b. In the Name column of the Data Sources table, click EssDS, EssInternalDS, and EssXADS, one by one, to change their passwords. The General tab of the Settings for EssDS/EssInternalDS/EssXADS is displayed.

    c. Select the **Connection Pool** tab.

    d. Enter and confirm the new password. All the three data sources - EssDS, EssInternalDS, and EssXADS - should have the same password, which is the `search_ess` schema password.

    e. Save the changes.

4. Restart the WebLogic Server.

Setting the SEARCH_MDS Schema User Password in WebLogic Server Middle Tier

If you change the password for the `search_mds` schema user of the Oracle SES database, then the data source connection pool password related to WebLogic Server configuration also needs to be changed.

> **Note:** If Oracle SES software is installed on an existing Database, then shut down the Oracle SES server and Oracle ESS server before executing the steps for changing the password related to the WebLogic Server configuration. Refer to "Starting and Stopping Oracle SES Instance" on page 2-6 for more information about starting and stopping WebLogic Server, Oracle ESS Server, and Oracle SES Server.

To change the data source connection pool password for SEARCH_MDS user:

1. Log in to the WebLogic Server Administration Console.

2. Under the **Change Center** in the left panel, click **Lock & Edit**.

3. Change the passwords for **mds-ESS_MDS_DS** and **mds-owsm** data sources:

   a. Under the **Domain Structure** in the left panel, select **search_domain > Services > Data Sources**. The Configuration tabs for the settings for mds-ESS_MDS_DS and mds-owsm is displayed in the main panel.

   b. In the Name column of the Data Sources table, click mds-ESS_MDS_DS and mds-owsm, one by one, to change their passwords. The General tab of the settings for mds-ESS_MDS_DS/mds-owsm is displayed.

   c. Select the **Connection Pool** tab.

   d. Enter and confirm the new password. All the two data sources - mds-ESS_MDS_DS and mds-owsm - should have the same password, which is the search_mds schema password.

   e. Save the changes.

4. Restart the WebLogic Server.

Temporary Password

For added security, a temporary password feature is provided. You can enter login credentials for use by the crawler when creating table sources, e-mail, OracleAS Portal, or Web sources. For Web sources, authentication can be performed with HTTP authentication, HTML forms, and OracleAS Single Sign-On.

To use the temporary password feature, select the **Delete Passwords After Crawl** option in the Oracle SES Administration GUI when creating or editing a source. This option is not available when self service for Web sources is enabled.

When a source has the **Delete Passwords after Crawl** option enabled, then you are prompted for all required passwords whenever the schedule for that source starts. You must start these schedules manually, which enables you to respond to the prompts. The supplied passwords are removed immediately after the schedule completes.

Authentication and Authorization

Oracle SES security is implemented at two levels: user authentication and user authorization.

This section contains the following topics:

- About Oracle SES Authentication
- Restrictions on Changing the ACL Policy
- Activating an Identity Plug-in

- [Re-registering Pre-Installed Identity Plug-ins](#)

- [Restrictions on Changing the Identity Plug-in](#)

About Oracle SES Authentication

User authentication identifies a user through an identity management system. You can register an identity plug-in to any identity management system; Oracle SES provides registered identity plug-ins for many identity management systems. The plug-in that you activate is responsible for all authentication and validation activity in Oracle SES. Activation is performed on the Global Settings - Identity Management Setup page.

Security filter configuration for the identity plug-in is performed on the Global Settings - Query Configuration page. A login does not force a refresh of the user's security filter. For a query request, Oracle SES checks the timestamp of an existing cached security filter and refreshes it when the specified life span has expired. The default latency is 60 minutes.

> **See Also:** "Activating an Identity Plug-in" on page 11-9 and "Identity Plug-in API" on page 13-48

About Oracle SES User Authorization

User authorization determines whether a user can access information about a particular item in the result list. It can be implemented in two layers.

The first layer utilizes access control lists (ACLs). An ACL lists the users or groups of users that have access to the document. The ACL can be assigned by the administrator to an entire source through the Oracle SES Administration GUI (*source-level ACLs*), or it can be provided by the source itself for each document (*document-level ACLs*).

The second layer uses a Java class to dynamically filter documents at search time (*query-time authorization*).

Oracle SES can use the following types of ACL policies:

- **Source-level ACLs**: An individual source can be protected by a single ACL, which governs access to every document in that source. These ACLs are defined on the Home - Sources - Authorization page.

- **Document-level ACLs**: Oracle SES provides mapped security to repositories by retrieving the ACL for each document at the time of crawling and indexing. At crawl time, the ACL for each document is passed to the crawler along with the document content, and the ACL is stored in the index. Currently Oracle SES supports document-level ACLs for user-defined sources and OracleAS Portal sources. (The ACL policy is Documents Controlled by the Source.) With user-defined sources, ACLs are returned by the crawler plug-in implemented by the user. With OracleAS Portal sources, ACLs are returned by the OracleAS Portal server. At search time, Oracle SES does not need any connection with the repository to validate access privileges.

For both source-level ACLs and document-level ACLs, all users and roles defined in the ACLs must exist in the identity plug-in.

User names are not case sensitive.

Table 11–1 identifies when documents are visible with the document ACL types supported in Oracle SES:

Table 11–1   Document ACL Types in Oracle SES Security Model

Document ACL Type	Public User	Authenticated User	Authenticated User with Allow Permission to Document	Authenticated User with Deny Permission to Document
No ACL	document visible	document visible	N/A	N/A
Deny Permission Only	--	--	N/A	--
Allow Permission Only	--	--	document visible	N/A
Deny with Allow Permissions	--	--	document visible	--

Table 11–2 compares the document-level user authorization methods in Oracle SES.

Table 11–2   User Authorization Methods in Oracle Secure Enterprise Search

Method	How Authorization is Determined	Advantages	Disadvantages
ACLs	The ACL is supplied by a crawler plug-in or an OracleAS Portal server.	Faster secure search performance. No additional programming is required for ACL-based OracleAS Portal security. (If implementing a crawler plug-in, then some additional work is necessary to supply ACLs.)	ACLs are static: they are updated only when crawling the source repository or when the administrator changes Oracle SES ACLs in the Oracle SES Administration GUI
Query-time Authorization	`ResultFilterPlugin` Java class.	Dynamic authorization. Reflects real-time user access privilege on documents.	There is performance overhead in cases when the search is not selective, returning large number of rows before query-time authorization. Extra work is required to implement a `ResultFilterPlugin`.

For sources that do not fit the user/group model, an authorization plug-in provides a more flexible security model. With an authorization plug-in, a crawler plug-in can add security attributes similar to document attributes. The authorization plug-in is invoked at login time to build security filters onto the query string. The security filters are applied against the values of the security attributes for each document. Only documents whose security attribute values match the security filter are returned to the user.

See Also:

- "Authorization Plug-in API" on page 13-48
- "Administrator-Based Authorization" on page 11-18 for more information about ACLs
- "Query-time Authorization" on page 11-20 for more information on Java filter classes
- "Crawler Plug-in API" on page 13-40

Restrictions on Changing the ACL Policy

On the Home - Sources - Authorization page, you can set and change the ACL policy.

The following ACL policy options are available:

- **No ACL**: With this setting, all documents are considered searchable and visible

- **Oracle Secure Enterprise Search ACL**: With this setting (also known as **source-level ACLs**), you can protect the entire source with one ACL. The same ACL protects every document in that source.

- **ACLs Controlled by the Source**: This setting (also known as **document-level ACLs**) is available only for OracleAS Portal sources and user-defined sources. This preserves authorizations specified in OracleAS Portal. For user-defined sources, crawler plug-ins (or connectors) can supply ACL information with documents for indexing, which provides finer control document protection. (That is, each document in the source can have different access privileges.)

The following restrictions apply to changing the ACL policy:

- If the schedule associated with the source is not currently being crawled, and if the source has never been crawled, then all ACL policy changes are allowed.

- If the schedule associated with that source is currently being crawled (that is, the schedule status is Launching, Executing, or Stopping), then all ACL options are grayed out, and you cannot change the ACL policy.

- If the schedule associated with the source is not currently being crawled, but the source *has* been crawled earlier, then the only change allowed is between **No ACL** and **Oracle Secure Enterprise Search ACL** (in either direction). This is visible in the Oracle SES Administration GUI as follows:

  – If the ACL option selected before the crawl started was **No ACL** or **Oracle Secure Enterprise Search ACL**, then the **ACLs Controlled by the Source** option is grayed out.

  If a secure ACL policy was selected but the identity plug-in is deactivated, then you can change the ACL policy to **No ACL** regardless of the crawl status.

- OracleAS Portal sources are subject to the same restrictions as other sources. That is, no changes are allowed while being crawled, and only changes between **No ACL** and **Oracle Secure Enterprise Search ACL** are allowed after crawling completes. However, the ACL policy for an OracleAS Portal source can also change if it is inheriting the ACL policy from its OracleAS Portal server parent; for example, when the OracleAS Portal server ACL policy is modified or when the OracleAS Portal source is changed from specifying the ACL policy locally to inheriting it from the server. Therefore, changes on an OracleAS Portal server are restricted so that no disallowed changes can occur on any children that inherit the ACL policy.

If any child inheriting the ACL policy is being crawled, then no changes are allowed on the OracleAS Portal server. If any child inheriting the ACL policy has been crawled, then the only changes allowed are between **No ACL** and **Oracle Secure Enterprise Search ACL**. (If the OracleAS Portal server policy is **ACLs Controlled by the Source**, then no changes are allowed). Similarly, the OracleAS Portal source cannot be set to inherit its ACL policy from the OracleAS Portal server if the associated change in ACL policy would be disallowed.

> **Note:** A source that is being crawled is different from a source whose associated schedule is being crawled. Oracle SES restricts all ACL policy changes for a source when the schedule associated with that source is being crawled. A source might not be crawled, but the schedule associated with it could be crawled if another source in the same schedule is being crawled.

Activating an Identity Plug-in

You can register an identity plug-in to any identity management system. Oracle SES provides registered identity plug-ins for many identity management systems. The plug-in that you activate is responsible for all authentication and validation activity in Oracle SES. Activate an identity plug-in on the Global Settings - Identity Management Setup page.

When you activate a plug-in, the Global Settings - Activate Identity Plug-in page is displayed. When completing this page, ensure that the values for Additional User Base and Additional Group Base are not subsets of Directory Subscriber.

> **Caution:** You must create the identity plug-in before creating an Oracle SES data source for any type of secure data. If the identity plug-in is not active, then the data source is crawled as a public data source.

The following table lists which identity plug-ins are available for each enterprise content source.

Table 11–3   Identity Plug-ins for Enterprise Content Sources

Source Type	Versions Supported	Identity Plug-in
Database	Any databases with a JDBC driver	Native
EMC Documentum Content Server	5.1, 5.2.5, 5.3 SP2	Active Directory, Oracle Internet Directory, Native
Lotus Notes	5.0.9, 6.5.4,7.0	Active Directory, Oracle Internet Directory, Native
Microsoft Exchange	Windows 2000, Windows 2003	Active Directory
Microsoft SharePoint Portal Server	2003, 2007	Active Directory
NTFS	Windows 2000, Windows 2003	Active Directory, Oracle Internet Directory
Oracle Calendar	10.1.2 or later	Oracle Internet Directory

Table 11–3   (Cont.) Identity Plug-ins for Enterprise Content Sources

Source Type	Versions Supported	Identity Plug-in
Oracle Content Database	Oracle Content Services 10.1.2 or later, Oracle Content Database 10.2 or 10.1.3	Native, Query-time authorization
Oracle E-Business Suite	11, 12	Native
Oracle Mail	10g	Oracle Internet Directory
Siebel 7.8	7.8	Native
Siebel 8	8	Native
Oracle Content Server	7.1.1, 7.5.2, 10gR3	Native
Oracle WebCenter	11g	Active Directory, Oracle Internet Directory, Sun Java System Directory Server for using Oracle Unified Directory (OUD)

Activating the Active Directory Identity Plug-in When connecting to Active Directory, Oracle SES tries to resolve the Active Directory domain name to an IP address of the Active Directory Domain Controller. This is generally not possible, especially when Oracle SES is installed on a non-Windows system or on a Windows system in a different domain. You must add the IP address of the Active Directory domain to the host file.

For example, to connect to an Active Directory domain called `foobar.example.com`, you must add something similar to the hosts file: `10.123.1.2 foobar.example.com`. Search for the hosts file in `C:\Windows\System32\Drivers\etc\HOSTS` on Windows systems, and `/etc/hosts` on UNIX systems.

For the Active Directory identity plug-in enter values for the following parameters:

- **Directory URL**: `ldap://`*`ActiveDirectoryserver`*`:389`

- **Directory account name**: *`UserLogonName`* Confirm the user logon name on the Active Directory Users and Computers application. Under the **User** folder, right-click **username**. Select **Property** and go to the **Account** tab. For example, assume the user account `adtest` in domain `domain1.example.com`, which is associated with the target Active Directory. You may try `domain1\adtest` or `adtest@domain1.example.com` or `cn=adtest,cn=users,dc=domain1,dc=example,dc=com` if you are not sure the actual user logon name. The user account does not need to be an administrator account.

- **Directory account password**: *`PasswordForDirectoryAccount`*

- **Directory subscriber**: dc=domain1,dc=example,dc=com for the domain name `domain1.example.com`

- **Directory security protocol**: none

If you deactivate an identity plug-in, then you must restart the middle tier.

Re-registering Pre-Installed Identity Plug-ins

If a pre-installed identity plug-in is accidentally removed, you can re-register it with the following steps:

1. On the Global Settings - Identity Management Setup page, click **Register New Identity Plug-in**.

2. Enter the class name and jar file name of the removed identity plug-in, as identified in Table 11–4.

3. Click **Finish**.

Table 11–4   Identity Plug-in Class Names and Jar File Names

Identity Plug-in	Plug-in Class Name	Jar File Name
Documentum Content Services	oracle.search.plugin.security.identity.dcs. DCSIdentityPluginManager	../dcs/DCSIdentityPlugin.jar
Database	oracle.search.plugin.security.identity.db. DBIdentityPluginManager	../oracleapplications/DBCrawler.jar
Oracle E-Business Suite	oracle.search.plugin.security.identity.ebs. EBSIdentityPluginMgr	../oracleapplications/EBSCrawler.jar
Oracle Fusion	oracle.search.plugin.security.identity.fusi on.FusionIdentityPluginMgr	../oracleapplications/FusionCrawler.jar
PeopleSoft	oracle.search.plugin.security.identity.psft. PsftIdentityPluginMgr	../oracleapplications/PeoplesoftCrawler. jar
Siebel 7.8	oracle.search.plugin.security.identity.sieb el.Siebel78IdentityPluginMgr	../oracleapplications/Siebel78Crawler.jar
Siebel 8	oracle.search.plugin.security.identity.sieb el.SiebelIdentityPluginMgr	../oracleapplications/Siebel8Crawler.jar
Oracle Content Server	oracle.search.plugin.security.identity.stell ent.StellentIdentityPluginMgr	../oracleapplications/StellentCrawler.jar
Oracle Internet Directory	oracle.search.plugin.security.identity.oid. OIDPluginManager	OIDPlugins.jar
Active Directory	oracle.search.plugin.security.idm.Identity PluginManagerADImpl	idm/idmPlugin.jar
Sun Java System Directory Server	oracle.search.plugin.security.idm.Identity PluginManagerIPlanetImpl	idm/idmPlugin.jar
OpenLDAP Directory	oracle.search.plugin.security.idm.Identity PluginManagerOpenLdapImpl	idm/idmPlugin.jar
Lotus Notes	oracle.search.plugin.security.identity.ln.L NIdentityPluginManager	ln/LNIdentityPlugin.jar

Restrictions on Changing the Identity Plug-in

The information Oracle SES saves from the identity plug-in (that is, the correspondence between names and canonical attribute values) may not be valid on different identity plug-ins. If you keep the same identity plug-in server (for example, to change port numbers or to switch to SSL), then you can deactivate and re-activate the identity plug-in anytime without restriction. This section describes steps you must perform if you change identity plug-in servers with user information that is not identical.

If you have sources using the ACL policy Oracle Secure Enterprise Search ACL and you decide to use a different identity plug-in server, then you must clear the ACL data before deactivating the original identity plug-in. If the ACL data is not cleared, then the ACL policy configured for that source while connected to the old identity plug-in server is not correctly enforced after connecting to the new identity plug-in server.

The existing ACL data can be cleared using either of these methods:

- Before deactivating the identity plug-in, for each source using the ACL policy Oracle Secure Enterprise Search ACL, switch the ACL policy to No ACL. After

connecting to the new identity plug-in server, restore the ACL policy to Oracle Secure Enterprise Search ACL and add the ACLs again. This temporarily makes the source public. If this is unacceptable, then use the next option.

- Before deactivating the identity plug-in, delete each source that uses the ACL policy Oracle Secure Enterprise Search ACL. After connecting to the new identity plug-in server, add the sources back and configure them again. The documents are never made public; but this may involve more work than the previous option.

If you have sources using the ACL policy ACLs Controlled by the Source and you decide to use a different identity plug-in server, then after activating the new identity plug-in server, each source that uses this ACL policy must be re-crawled with the **Process All Documents** option. This forces the reloading and indexing of all of ACL information for such sources using the new identity plug-in server. Select the **Process All Documents** option on the Home - Schedules - Edit Schedule page.

> **Note:** If the ACL data is not cleared before switching identity plug-in servers, then you see a message that some users and groups could not be found by the identity plug-in. Those users and groups are still displayed on the Home - Sources -Authorization page. They can be deleted manually.

Authentication Methods

The Oracle SES front-end interface collects user credentials, which are then validated against the active identity plug-in. In addition to authentication of search users, Oracle SES must also authenticate the crawler when accessing external data repositories. Administrators supply credentials to crawl private content through the following authentication methods:

- HTTP authentication (both basic and digest authentication)
- HTML forms
- OracleAS Single Sign-On
- Oracle Access Manager (OAM) Single Sign-On

It is the administrator's responsibility to check the authorization policy to ensure that crawled documents are properly protected.

Oracle Secure Enterprise Search User Repository

Oracle SES has two types of users:

- **Administrative User**: The administrative user is searchsys by default. This user is natively defined in Oracle SES. Only this user can use the Oracle SES Administration GUI.

  If the Single-Sign On (SSO) is enabled, then the SSO user can also use the Oracle SES Administration GUI.

- **Search Users**: Oracle SES lets you register an identity plug-in as an interface to any identity management system. (Oracle SES provides registered identity plug-ins for Oracle Internet Directory and other identity management systems.) The plug-in that you activate is responsible for all authentication and validation activity in Oracle SES. Use the Global Settings - Identity Management Setup page in the Oracle SES Administration GUI to associate Oracle SES with an identity management system.

Oracle Internet Directory is Oracle's native LDAP v3-compliant directory service. It is part of the Oracle Identity Management infrastructure. It is not included in Oracle SES. Oracle Internet Directory should be version 9.0.4 or 10.1.2 (with the latest patch release applied) for connection with Oracle SES. Oracle Internet Directory is not a part of Oracle SES, and therefore Oracle SES can be linked to any existing or new Oracle Internet Directory.

Oracle Secure Enterprise Search Authentication Interface

For the `searchsys` administrative user, the login screen is available in the Oracle SES Administration GUI. If the Single Sign-on (SSO) is enabled, then the administrative user can access the Oracle SES Administration GUI using the SSO login screen.

For search users, there are three possible front-end authentication interfaces:

- HTML form login page. Oracle SES provides an authentication page, and it authenticates against the identity plug-in.

- Web Services API. The `login` and `logout` Web Services operations authenticate against the identity plug-in.

- Single Sign-on login screen. This can be made available by using Oracle SES with OracleAS Single Sign-On and Oracle HTTP Server. These are available as part of the Oracle Identity Management infrastructure in OracleAS.

> **Note:**
>
> - At any point in time, only one method of authentication can be used for a search user or for an administrative user.
>
> - Oracle strongly recommends that you SSL-protect the channel between the Oracle HTTP Server and the Oracle WebLogic server instance for secure content.

Enabling Secure Search

Much of the information within an organization is publicly accessible. However, there are other sources that are protected. For example, while a user can search in their own e-mail folders, they should not be able to search anyone else's e-mail. A secure search returns only search results that the user is allowed to view based on access privileges.

Oracle SES can use the following two security modes: OracleAS Single Sign-on (SSO) or non-SSO. This is set on the Global Settings - Query Configuration page:

- Require login to search secure content only: anyone can search public content. This is the default. This is also known as secure mode 2.

- Require login to search secure *and* public content. This is also known as secure mode 3.

This is applied to both the default query application and Oracle SES Web services. In mode 3, if a user tries to perform any Web services operation (search or document service) without logging in first, then a SOAP exception is thrown indicating that this secure mode requires login for any operation.

This section describes the authorization methods that Oracle SES supports. The authorization methods prevent search users from accessing documents for which they do not have privileges.

Oracle Secure Enterprise Search offers several options for secure search:

- [User Authorization Cache](#)

- [Federated User Authorization Cache](#)

- [Administrator-Based Authorization](#)

- [Custom Crawler Plug-in](#)

- [Identity-Based Secure Search](#)

- [Query-time Authorization](#)

- [Self Service Authorization](#)

> **See Also:** *Oracle SES Administration Tutorial* at
> `http://www.oracle.com/pls/topic/lookup?ctx=E35215-01&id=tut_`
> `ses_administration`

User Authorization Cache

The User Authorization Cache (UAC) source type can crawl and cache user authorization information such as groups and accessible values of user security attributes. This cached information is used at query time to build a security filter. Querying a local cache is much faster than retrieving the authorization information from external repository and identity systems, and thus it significantly reduces the time to build the security filters for the current user. As a result, users can log in to Oracle SES much more quickly. Moreover, you can set up UAC sources to crawl the user authorization information off line, which reduces the load on target repositories at query time.

You can use UAC for sources that are based on either of the security models supported in Oracle SES:

- **Identity-based security**: UAC is enabled for Oracle Internet Directory, Active Directory, and Lotus Notes.

- **Attribute-based security**: UAC is enabled for Oracle Content Database and Oracle Content Server sources. This type of security is also called the user-defined security model.

The crawler stores the following information in the User Authorization Cache:

- User groups: The list of groups that a user belongs to.

- User attribute values: The values of a specified list of attributes for particular data sources. The values can be single values or arrays of values.

To create a UAC source for an identity plug-in:

1. Click the **Sources** tab.

2. For Source Type, select **User Authorization Cache**, then click **Create**.

   The Create User-Defined Source page is displayed.

3. Configure the UAC source with the parameters described in Table 11–5. Set **Retrieve user groups** to true.

4. Create and activate the identity plug-in. Configure the plug-in to use the cache.

   For example, set these parameters for an OID plug-in:

   - **Use User Cache**: `true`

   - **User Cache Source Name**: Name of the UAC source that caches the user group information.

To create a UAC source for attribute-based security:

1. Click the **Sources** tab.

2. For Source Type, select **User Authorization Cache**, then click **Create**.

   The Create User-Defined Source page is displayed.

3. Configure the UAC source with the parameters described in Table 11–5. Be sure to set **Source names for which security attributes should be crawled**.

4. Configure the data sources to use cached user authorization information for building security filter by setting appropriate UAC related parameters in the authorization plug-in.

   For example, set these parameters for Oracle Content Server source:

   - **Use cached user and role information to authorize results**: true

   - **User role data source to cache the filter**: Name of the UAC source that caches the user authorization information for this data source.

Table 11–5   User Authorization Cache Parameters

Parameter	Setting
User search query	Query expression defining the set of users to be crawled. For example, enter a* to crawl all users whose names begin with the letter a. Leave this parameter blank to crawl all users who have logged into Oracle SES. This expression is interpreted by the active identity plug-in. When the active identity plug-in is Oracle Internet Directory or Active Directory, specify the LDAP query to select the set of users to crawl.
User attributes to be synchronized from identity system	Comma-delimited list of user attributes to crawl, such as `name`, `groups`. Not used in this release: Leave blank.
Retrieve user groups	Enter `true` to cache groups for users; otherwise, enter `false`.
Source names for which security attributes should be crawled	Enter a comma-delimited list of source names with security attributes to crawl. For example, `source1, source2`. The security model in these sources must be attribute-based security.

Federated User Authorization Cache

The Federated User Authorization Cache maintains a single User Authorization Cache (UAC) for use by all Oracle SES instances in a federated environment. Any identity plug-in or authorization plug-in can use a Federated UAC.

Prerequisite

- Define one or more UAC sources for an Oracle SES instance in the federated environment, as described in "User Authorization Cache" on page 11-14. A local cache created on this system can be accessed as a federated UAC source by any other instance with an identically configured identity or authorization plug-in.

To create a federated UAC:

1. Click the **Sources** tab.

2. For Source Type, select **Federated UAC**, then click **Create**.

   The Create Federated UAC page is displayed.

3. Configure the federated UAC source with the parameters described in Table 11–6.

4. Configure an identity or authorization plug-in:

    a. Select the **Global Settings** secondary tab.

    b. Under System, choose **Identity Management Setup**.

    c. Activate an identity or authorization plug-in. Enter the name of the federated UAC as the value of the **User Cache Source Name** parameter.

5. Repeat these steps for each additional Oracle SES instance in the federated environment.

Table 11–6   Federated UAC Parameters

Parameter	Setting
Connection String	JDBC connection string to a remote Oracle SES instance with one or more UAC sources. (Required)
Remote Cache Config File	Full path name of the XML configuration file for the remote cache, as described in Example 11–1. (Required)
	If the Oracle SES software is deployed across multiple middle tiers, then this remote cache configuration file must be accessible to all the middle tiers.
Password	Password for the searchsys user on the Oracle SES instance identified by **Connection String**. (Required)

Modifying the Remote Cache Configuration File

You can create the remote cache configuration file anywhere on the computer where you defined a federated UAC. You identify the location of the file when defining a Federated UAC source. The file provides details about the remote UAC source. Example 11–1 shows a sample file.

Example 11–1   Remote Cache Configuration File

```
<?xml version="1.0" encoding="UTF-8"?>
<FederatedUAC>
    <UserCache>
        <Name>UAC1</Name>
    </UserCache>
    <UserCache>
        <Name>UAC2</Name>
        <UserRouting>u6-u10</UserRouting>
    </UserCache>
    <UserCache>
        <Name>UAC3</Name>
        <SourceMapping>
            <RemoteSourceName>S5</RemoteSourceName>
            <LocalSourceName>S1</LocalSourceName>
        </SourceMapping>
        <SourceMapping>
            <RemoteSourceName>S7</RemoteSourceName>
            <LocalSourceName>S3</LocalSourceName>
        </SourceMapping>
    </UserCache>
</FederatedUAC>
```

Name

The name of a UAC source in a remote instance of Oracle SES from which the federated UAC retrieves user and attribute information. (Required)

UserRouting

Not supported in this release.

SourceMapping

Maps the local source to a remote source with an identically configured authorization plug-in. (Required for attribute-based security)

Use these elements to provide the mapping information:

■ **RemoteSourceName**: Name of a remote source defined with a UAC and an authorization plug-in that is configured identically to the plug-in for **LocalSourceName**. (Required)

■ **LocalSourceName**: Name of the local source defined with a federated UAC and an authorization plug-in. (Required)

XML Schema Definition for Remote Cache Configuration Files

The following is the XML schema definition for remote cache configuration files:

```
<?xml version="1.0" encoding="windows-1252" ?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            xmlns="http://www.example.org" attributeFormDefault="unqualified">
  <xsd:element name="FederatedUAC">
    <xsd:annotation>
      <xsd:documentation>Federated Source Configuration</xsd:documentation>
    </xsd:annotation>
    <xsd:complexType>
      <xsd:all>
        <xsd:element name="UserCache">
          <xsd:annotation>
            <xsd:documentation>
              Remote UAC Source Config Details
            </xsd:documentation>
          </xsd:annotation>
          <xsd:complexType>
            <xsd:all>
              <xsd:element name="Name" minOccurs="1">
                <xsd:annotation>
                  <xsd:documentation>
                    UAC source name in remote instance from which
                    user/attribute information must be retrieved
                  </xsd:documentation>
                </xsd:annotation>
              </xsd:element>
              <xsd:element name="UserRouting" minOccurs="0">
                <xsd:annotation>
                  <xsd:documentation>
                     Which users should be routed to this UAC cache. Used for
                     identity-based security model.
                  </xsd:documentation>
                </xsd:annotation>
              </xsd:element>
              <xsd:element name="SourceMapping" minOccurs="0" maxOccurs="1">
                <xsd:annotation>
                  <xsd:documentation>
                    To map data source prefix for user-defined security
                    attribute retrieval. This will be used in case of user
                    defined security model. The mapping information must be
                    mentioned in the form of remote and local data source names.
                  </xsd:documentation>
```

```
                                        </xsd:annotation>
                                <xsd:complexType>
                                  <xsd:all>
                                    <xsd:element name="RemoteSourceName" maxOccurs="1"
minOccurs="1">
                                        <xsd:annotation>
                                          <xsd:documentation>
                                            Remote Instance Data source name prefixed to the
                                            attribute while caching security attribute information
                                          </xsd:documentation>
                                        </xsd:annotation>
                                        <xsd:complexType/>
                                    </xsd:element>
                                    <xsd:element name="LocalSourceName" minOccurs="1">
                                      <xsd:annotation>
                                        <xsd:documentation>
                                          Local Instance Data source name for which security
                                          attribute is being retrieved
                                        </xsd:documentation>
                                      </xsd:annotation>
                                    </xsd:element>
                                  </xsd:all>
                                </xsd:complexType>
                              </xsd:element>
                            </xsd:all>
                          </xsd:complexType>
                      </xsd:element>
                  </xsd:all>
              </xsd:complexType>
          </xsd:element>
</xsd:schema>
```

Administrator-Based Authorization

With administrator-based authorization, the administrator can specify an
authorization policy when creating a source. This policy governs which users can view
each document. Administrator-based authorization is based on ACLs. When a source
is crawled, each document is stamped with an ACL. When a user enters a search, the
result list only includes documents for which the user credentials match the document
ACL.

> **See Also:**   "Authentication and Authorization" on page 11-5 for more
> information about ACL policies

Oracle SES performs ACL duplicate detection. If a crawled document's ACL exists in
the Oracle SES system, then that ACL is used to protect the document, instead of
creating a new ACL within Oracle SES. This policy reduces storage space and
increases performance.

Oracle SES supports only a single LDAP domain. The LDAP users and groups
specified in the ACL must belong to the same LDAP domain.

> **Caution:** If ACLs are crawled from sources, then ensure that the sources being crawled belong to the same LDAP domain. Otherwise, the end users might inadvertently be granted permission to documents that they should not be able to access.
>
> When secure search is enabled, you may encounter up to a 15 minute delay viewing the private documents. This delay could be due to newly added secure sources or a user/group membership change in the identity management system.

Custom Crawler Plug-in

Oracle Secure Enterprise Search provides an API for writing custom crawler plug-ins (or **connectors**) in Java. With this API, you can create a secure crawler plug-in to meet your requirements.

The custom crawler plug-in passes back URLs directly to be indexed. Each URL can be accompanied by an ACL, which restricts the access to that particular document. Alternatively, an ACL can be set in the Oracle SES Administration GUI for the source.

Authentication credentials can be provided to the plug-in through parameters in the Oracle SES Administration GUI. The plug-in uses these credentials to access the secure source.

Within the Crawler Plug-in API, the `DocumentAcl` object implements **identity-based security**. Identity-based security is a security policy based on users and groups that is defined by the active identity plug-in.

> **See Also:** "Crawler Plug-in API" on page 13-40

Identity-Based Secure Search

You can do identity-based secure search with administrator-based authorization or custom crawler plug-ins.

Oracle SES provides identity plug-ins for OpenLDAP release 2.2 and 2.3 and Sun Java System Directory Server release 5.1 and 5.2. Activate either of these identity plug-ins on the Global Settings - Identity Management Setup page.

To use administrator-based authorization:

1. On the Home - Sources page, select a source to use administrator-based authorization.

2. On the Home - Sources - Customize Source page, click the **Authorization** tab.

3. Under Crawl-time ACL Stamping, select **Oracle Secure Enterprise Search ACL**.

4. For Type, select **User** or **Group**.

5. Click **Add Another Row**.

6. For **User**, select **USER_NAME** or something you want to use as **Format** and enter user name as **Name**. For **Group**, select **DN** as **Format** and input **cn=<Group>,<Group search bases>** as **Name**.

7. Click **Apply**.

To use custom crawler plug-ins:

1. Create a custom crawler plug-in with the Crawler Plug-in API. See "Crawler Plug-in API" on page 13-40.

2. For User & Group, add the following line:

```
DocumentAcl acl;
// For User
acl.addPrincipal("<User>", DocumentAcl.SIMPLE, DocumentAcl.USER);
// For Group
acl.addPrincipal("cn=<Group>,<Group search bases>", DocumentAcl.DN,
DocumentAcl.GROUP);
```

3. If you get an error registering the identity plug-in, check the WebLogic log files at *wls_domain_home/ses_domain_name*/servers/AdminServer/logs/.

Limitations with OpenLDAP and Sun Java System Directory Identity Plug-ins

The LDAP entry of users and groups on OpenLDAP or Sun Java System Directory Server requires the following conditions:

Users:

- Belong to the following objectClasses: person, organizationalPerson, and inetOrgPerson

- Have the following attributes: dn, cn, sn

- The entry's location: uid=<User>,<User search bases>

Groups:

- Belong to one objectClasse: groupOfUniqueNames

- Have the following attributes: dn, uniqueMember

- The entry's location: cn=<Group>,<Group search bases>

Query-time Authorization

Query-time authorization provides another form of filtering. Query-time authorization can be enabled or disabled for Web, file, table, e-mail, mailing list, OracleAS Portal, and user-defined source types from the Home - Sources - Edit Source page. It is not available for federated or self-service sources. Query-time authorization can be used with or without ACLs. For example, a source could be stamped with a relatively broad ACL, while query-time authorization could be used to further filter the results.

In query-time authorization, the Oracle SES administrator associates a Java class that is called at run time. The Java class validates each document that is returned in a user query.

> **See Also:** "Query-time Authorization API" on page 13-49 for more information about the Query-time Authorization API of Oracle SES Java SDK

Query-time authorization requires these steps:

1. The Oracle SES administrator registers a Java class implementing the ResultFilterPlugin interface with a source that requires query-time authorization.

2. Oracle SES crawls, collects, and indexes all documents. If ACL stamping has been set up, then it ACL stamps the documents.

3. At search time, the search result list initially contains all documents accessible under crawl-time ACL policies, unfiltered by query-time user privilege checking.

4. For the top-N results requested by the user, Oracle SES calls the registered Java class, passing in the search request and document information for any documents belonging to the protected source. The Java class returns an integer value for each document indicating if the document should be removed from the result or not.

5. Only items the user is privileged to see are returned to the user in their result list.

Notes for Using Query-time Authorization:

- The Browse application is also filtered by the query-time authorization mechanism. The `ResultFilterPlugin` class controls which folders are visible to the user, and documents within folders are filtered by the same process as the standard search result list.

- Set the **Hit Count Method** to **Exact count (adjusted for query-time filtering)** on the Global Settings - Query Configuration page. If it is not set to **Exact Count**, then the hit count displayed could be larger than the actual number of documents the user is authorized to view. The page in the Oracle SES Administration GUI contains other query-time authorization configuration settings you might want to consider.

  Oracle SES reports an approximate count of search results. The number of documents Oracle SES fetches determines the accuracy of the estimation. When the hit count is high enough to go beyond one page of results, then the count changes to a more accurate count as you click **Next** pages. **Exact count** shows an accurate count, but this option impacts query response time.

  > **Note:** While crawling secure data sources, the crawler skips **infosource** count-hit. Changing the ACL policy for a data source from secure to public does not result in an automatic update of the infosource count-hit. If you wish to include the data source in the infosource count-hit, you must recrawl the data source or trigger infosource `update_doc_count` for the data source.

- If you modify the contents of the jar file containing the `ResultFilterPlugin` implementation classes, but do not change the location of the jar file, then you must restart the Oracle SES middle tier. This ensures that the query application picks up your changes and that the Java Virtual Machine does not use a cached version of the class within the old jar file. Restart the middle tier.

- If a `ResultFilterPlugin` class is enabled for an OracleAS Portal server, then all of its page group sources are automatically protected by that query-time filter.

- It may take up to five seconds for query-time authorization changes applied in the Oracle SES Administration GUI to take effect in the Oracle SES search engine. The relevant settings are the following:

  - Enabling a `ResultFilterPlugin` class for a source

  - The hit count method

  - The Query-time Authorization Configuration settings on the Global Settings - Query Configuration page.

Self Service Authorization

Self service authorization allows end users to enter their credentials needed to access an external content repository. Oracle Secure Enterprise Search crawls and indexes the repository using these credentials to authenticate as the end user. Only the self service

user is authorized to see these documents in their search results. Self service authorization works well out of the box, as the crawler appears to be a normally authenticated end user to the content repository.

To set up a self service source, create a template source, defining the target data repository but omitting the credentials needed to crawl. From the query application, an end user can view the Customize page and subscribe to a template source by entering their credentials in an input form. A new user-subscribed source is created, along with a copy of the template's schedule. Oracle SES creates an ACL for this user to be applied to the source.

User-subscribed sources are viewable in the Home - Sources - Manage Template Source page, and the associated schedules are administered in the Home - Schedules page. Any changes applied by the administrator to a template source are dynamically inherited by the associated user-subscribed sources for the next crawl.

The self service option is available for e-mail and Web sources. Self service e-mail sources require the administrator to specify the IMAP server address and the end user to specify the IMAP account user name and password. Self service Web sources are limited to content repositories that use OracleAS Single Sign-On authentication. The administrator specifies the seed URLs, boundary rules, document types, attribute mappings, and crawling parameters, and the end user specifies the single sign-on user name and password.

Crawling of user-subscribed sources is controlled by the administrator. End users do not see any search results for their subscribed source until that source is crawled by the administrator's schedule. Allowing a crawl to automatically launch immediately after an end user subscribes to a source might be useful. However, it makes it possible for users to load the system at inconvenient times.

Configuring Secure Search with OracleAS Single Sign-On

If you use OracleAS Single Sign-On (SSO), then you can configure Oracle SES to use your SSO server for authentication. This section describes the necessary configuration steps.

OracleAS supported version with the current Oracle SES installation is 10.1.4.0.1, with the latest patch sets applied. The supported Oracle HTTP Server versions are 10.1.3 and 11g.

> **Note:** `ORACLEAS_HOME` refers to the Oracle home directory of the OracleAS middle tier installation, which is typically stored in a system variable. On UNIX systems, you can reference the path as `$AS_HOME`. On Windows, the equivalent is `%AS_HOME%`.
>
> `ORACLEOHS_HOME` refers to the home directory of the Oracle HTTP server installation.

To enable Oracle SES for OracleAS Single Sign-on:

1. Front the Oracle SES instance with Oracle HTTP Server, as described in "Configuring Oracle HTTP Server".

2. Configure OracleAS and Oracle SES, as described in "Configuring OracleAS and Oracle SES for Single Sign-on Security" on page 11-27.

Configuring Oracle HTTP Server

The Oracle SES middle tier runs on the Oracle WebLogic Server, which is installed in *wls_home* directory.

You can configure either Oracle HTTP Server 10.1.3 or Oracle HTTP Server 11g. Special configuration is necessary on both the Oracle SES side and the Oracle HTTP Server (OHS) side.

> **Note:** When using Oracle HTTP Server fronting, Oracle SES allows the Oracle HTTP Server to assert the identity of the current user. You must limit this privilege to only trusted Oracle HTTP Server instances by SSL-protecting the communication between Oracle SES and Oracle HTTP Server.

To configure Oracle HTTP Server 10.1.3:

1. Copy the file `mod_wl_20.so` from *wls_home*`/server/plugin/linux/i686` to *ORACLEOHS_HOME*`/ohs/modules`.

2. Edit the file `httpd.conf` available at *ORACLEOHS_HOME*`/ohs/conf/` to include the following lines:

```
LoadModule weblogic_module [Complete path to mod_wl_20.so]
<IfModule mod_weblogic.c>
   WebLogicHost [SES host name]
   WebLogicPort [SES HTTP port]
</IfModule>
<Location /search/query>
   SetHandler weblogic-handler
</Location>
<Location /search/admin>
   SetHandler weblogic-handler
</Location>
```

For example, if the ORACLEOHS_HOME is `/ohsHome`, Oracle SES host is `sesHost` and the Oracle SES port is `8001`, then the file contains the following content:

```
LoadModule weblogic_module  /ohsHome/ohs/modules/mod_wl_20.so
<IfModule mod_weblogic.c>
   WebLogicHost sesHost
   WebLogicPort 8001
</IfModule>
<Location /search/query>
   SetHandler weblogic-handler
</Location>
<Location /search/admin>
   SetHandler weblogic-handler
</Location>
```

3. Edit the `httpd.conf` to comment out or remove the following lines:

```
#LoadModule auth_module modules/mod_auth.so
#LoadModule auth_anon_module modules/mod_auth_anon.so
#LoadModule auth_db_module modules/mod_auth_dbm.so
```

4. Register Oracle HTTP Server `mod_osso` with Oracle Single Sign-On server 10.1.4. To do register it, run the following command from *ORACLEAS_HOME*:

```
>setenv ORACLE_HOME [full path to $OracleAS_Home]
```

```
>$OracleAS_Home/sso/bin/ssoreg.sh -oracle_home_path  [Complete Path of
OracleAS_Home]  -site_name  [Hostname Of The Fronting OHS Server:Port]
-config_mod_osso TRUE  -mod_osso_url [http:// Hostname Of The Fronting OHS
Server:Port] -update_mode  CREATE  -remote_midtier  -config_file [Path to
generate the file osso.conf]
```

For example, if you installed Oracle Identity Management 10g in /asHome, and the Oracle HTTP Server URL is http://ohsserver:7779, then you must run the following command:

```
>setenv ORACLE_HOME /asHome
```

```
>$OracleAS_Home/sso/bin./ssoreg.sh -oracle_home_path /asHome -site_name
ohsserver:7779 -config_mod_osso TRUE -mod_osso_url http://ohsserver:7779
-update_mode CREATE -remote_midtier -config_file /temp/osso.conf
```

5. Configure mod_osso to protect web resources with static directives. Take these steps:

   ■ Copy the file osso.conf generated in step 4 to: *ORACLEOHS_HOME*/ohs/conf/osso.

   ■ Edit the file mod_osso.conf available at *ORACLEOHS_HOME*/ohs/conf/ to include the following:

```
LoadModule osso_module [path of OracleOHS_Home]/ohs/modules/mod_osso.so
<IfModule mod_osso.c>
  OssoIdleTimeout off
   OssoIpCheck on
   OssoConfigFile [path of OracleOHS_Home]/ohs/conf/osso/osso.conf
#OssoRedirectByForm off
#OssoSecureCookies on
#OssoProtectedOnly on
#OssoSecureCookies on
#OssoSendCacheHeaders on
#OssoHttpsFrontend on
#UseWebCacheIp on
<Location /search/query/formlogin.uix>
   require valid-user
   AuthType Basic
</Location>
</IfModule>
```

6. Edit the file httpd.conf to include the following:

```
include OracleOHS_Home/ohs/conf/mod_osso.conf
```

For example, if your ORACLEOHS_HOME is /ohsHome, then the value would be:

```
include /ohsHome/ohs/conf/mod_osso.conf
```

7. Restart the HTTP server:

```
> $OracleOHS_Home/opmn/bin/opmnctl restartproc process-type=HTTP_Server
```

To configure Oracle HTTP Server 11g:

1. Set *ORACLE_HOME* and *ORACLE_INSTANCE*.

   For example, if Oracle HTTP Server is installed at /ohsHome:

```
>setenv ORACLE_HOME /ohsHome
```

If the instance is named `myInstance`

```
>setenv ORACLE_INSTANCE /ohsHome/instances/myInstance
```

2. Edit the file `mod_wl_ohs.conf` available at *ORACLEOHS_HOME*/instances/instance1/config/OHS/ohs1/ to include the following:

```
<IfModule weblogic_module>
  WebLogicHost [SES host name]
  WebLogicPort [SES HTTP port]
  Debug ON
  WLLogFile Convenient Location of the log
</IfModule>
<Location /search/query>
  SetHandler weblogic-handler
</Location>
<Location /search/admin>
  SetHandler weblogic-handler
</Location>
# For monitor SES URL
<Location /monitor>
SetHandler weblogic-handler
</Location>
# For Help links in Admin side
<Location /search/ohw>
SetHandler weblogic-handler
</Location>
```

For example, if your Oracle SES host is `sesHost` and Oracle SES port is `8001`,

```
<IfModule weblogic_module>
   WebLogicHost sesHost
   WebLogicPort 8001
   WLLogFile /scratch/exampleuser/weblogic.log
</IfModule>
<Location /search/query>
   SetHandler weblogic-handler
</Location>
<Location /search/admin>
   SetHandler weblogic-handler
</Location>
<Location /monitor>
   SetHandler weblogic-handler
</Location>
<Location /search/ohw>
   SetHandler weblogic-handler
</Location>
```

3. Register Oracle HTTP Server `mod_osso` with Oracle Single Sign-On Server 10.1.4. Run the following command from *OracleAS_Home*:

```
>setenv ORACLE_HOME [full path to $OracleAS_Home]

>$OracleAS_Home/sso/bin/ssoreg.sh -oracle_home_path  [Complete Path of
OracleAS_Home]  -site_name [Hostname Of The Fronting OHS Server:Port]
-config_mod_osso TRUE  -mod_osso_url [http:// Hostname Of The Fronting OHS
Server:Port] -update_mode  CREATE  -remote_midtier -config_file [Path to
generate the file osso.conf]
```

For example, if you installed Oracle Identity Management 10g in /asHome, and the OHS URL is http://ohsserver:7779, then you need to run the following command:

```
>setenv ORACLE_HOME /asHome

>$OracleAS_Home/sso/bin./ssoreg.sh -oracle_home_path /asHome -site_name
ohsserver:7779 -config_mod_osso TRUE -mod_osso_url http://ohsserver:7779
-update_mode CREATE -remote_midtier -config_file /temp/osso.conf
```

4. Configure mod_osso to protect Web resources with static directives:

   a. Copy osso.conf generated above to the location *OracleOHS_HOME*/instances/instance1/config/OHS/ohs1/conf

   b. Copy the file mod_osso.conf from *OracleOHS_HOME*/instances/instance1/config/OHS/ohs1/disabled/ to OracleOHS_HOME/instances/instance1/config/OHS/ohs1/moduleconf/.

   c. Edit the file mod_osso.conf located at *OracleOHS_HOME*/instances/instance1/config/OHS/ohs1/moduleconf to include the following lines:

```
LoadModule osso_module ${OracleOHS_HOME}/ohs/modules/mod_osso.so
<IfModule mod_osso.c>
  OssoIdleTimeout off
  OssoIpCheck on
  OssoSecureCookies Off
OssoConfigFile  <path_to_osso.conf_file>
#Location is the URI you want to protect
<Location />
  require valid-user
  AuthType Osso
</Location>
</IfModule>
```

   For example if your OracleOHS_Home is /ohsHome and you generated the osso.conf file at *OracleOHS_HOME*/instances/instance1/config/OHS/ohs1/conf, then you need to include the following:

```
LoadModule osso_module "${OracleOHS_HOME}/ohs/modules/mod_osso.so"
<IfModule osso_module>
   OssoIdleTimeout off
   OssoIpCheck on
   OssoSecureCookies Off
   OssoConfigFile
/ohsHome/instances/instance1/config/OHS/ohs1/conf/osso.conf
<Location /search/query/formlogin.uix>
   require valid-user
   AuthType Osso
</Location>
</IfModule>
```

   d. Edit the file httpd.conf located at *OracleOHS_HOME*/instances/instance1/config/OHS/ohs1/ to include the following at the end of the file:

```
include "${ORACLE_INSTANCE}/config/${COMPONENT_TYPE}/${COMPONENT_NAME}/moduleconf/mod_osso.conf"

# Include the configuration files needed for mod_weblogic
```

```
include "${ORACLE_INSTANCE}/config/${COMPONENT_TYPE}/${COMPONENT_NAME}/mod_
wl_ohs.conf"

# Include the SSL definitions and Virtual Host container
include "${ORACLE_INSTANCE}/config/${COMPONENT_TYPE}/${COMPONENT_
NAME}/ssl.conf"
# Include the admin virtual host (Proxy Virtual Host) related configuration
include "${ORACLE_INSTANCE}/config/${COMPONENT_TYPE}/${COMPONENT_
NAME}/admin.conf"

include "moduleconf/*.conf"
```

5. Restart the HTTP server with the command:

   ```
   /$OracleOHS_Home/instances/instance1/bin/opmnctl startproc process-type=OHS
   ```

Configuring OracleAS and Oracle SES for Single Sign-on Security

To configure OracleAS SSO for Oracle SES:

1. Add OSSO provider for WebLogic domain as described in "Adding OSSO Provider for WebLogic Domain" on page 11-27.

2. Add OSSO Identity Asserter as described in "Adding OSSO Identity Asserter" on page 11-27.

3. Add OID Authenticator as described in "Adding OID Authenticator" on page 11-28.

4. Enable OSSO for Oracle SES as described in "Enabling OSSO for Oracle SES" on page 11-29.

Adding OSSO Provider for WebLogic Domain

To add an OSSO provider for WebLogic domain, perform the following steps:

1. Copy the file *ORACLEOHS_HOME*/modules/oracle.ossoiap_11.1.1/ossoiap.jar to the following WebLogic Server directory:

   wls_home/server/lib/mbeantypes

   > **Note:** The ossoiap.jar file is available only with OHS 11g. If you have configured SSO with OHS 10g, then you must separately install OHS 11g to obtain this file.

2. Restart the Oracle SES middle tier.

Adding OSSO Identity Asserter

To add an OSSO Identity Asserter to the domain, perform the following steps:

1. Log in to the WebLogic Server Administration Console.

2. Click **Security Realms**, *Default Realm Name*, **Providers**.

3. Click **New** under the Authentication Providers table.

4. Enter a name for the new provider, select its type, and then click **OK**. For example:

   Name: OSSO Identity Asserter

   Type: OSSOIdentityAsserter

5. Click the name of the newly added provider.

6. On the Common tab, set the appropriate values for common parameters and set the Control Flag to `SUFFICIENT`.

7. Save the changes.

8. Restart the WebLogic Server.

Adding OID Authenticator

To add an Oracle Internet Directory (OID) Authenticator to the domain, use the Oracle WebLogic Administration Console to perform the following steps:

1. Log in to the WebLogic Administration Console.

2. Click **Security Realms**, *Default Realm Name*, **Providers**.

3. Click **New** under the Authentication Providers table.

4. Enter a name for the new provider, select its type, and then click **OK**. For example:

   Name: `OID Authenticator`

   Type: `OracleInternetDirectoryAuthenticator`

5. Click **Save**.

6. Click the newly added authenticator to see the Settings page. Retain the default settings; do not change the Control Flag until you have verified that the Oracle Internet Directory configuration is valid.

7. On the Common tab, specify the following required settings and then save the settings.

   Propagate Cause For Login Exception: Check.

   Principal: LDAP administrative user. For example: cn=orcladmin

   Host: The Oracle Internet Directory host name

   Use Retrieved User Name as Principal: Check

   Credential: LDAP administrative user password. For example: password

   Confirm Credential: For example: password

   Group Base DN: Oracle Internet Directory group search base

   User Base DN: Oracle Internet Directory user search base.

   Port: Oracle Internet Directory port

8. Save all configuration settings.

9. Stop and restart the Oracle WebLogic Server for the changes to take effect.

After adding OracleAS Single Sign-On Identity Asserter and Oracle Internet Directory Authenticator as authentication providers, perform the following steps:

1. Log in to the WebLogic Administration Console.

2. Click **Security Realms**, **Default Realm Name**, **Providers**.

3. Select the **Users and Groups** tab to see the list of users and groups contained in the configured authentication providers.

   You should see user names from the Oracle Internet Directory configuration, which implicitly verifies that the configuration is working.

4. If the Oracle Internet Directory instance is configured successfully, you can change the control flag.

   If the Oracle Internet Directory authentication is sufficient for an application to identify the user, then choose the SUFFICIENT flag. SUFFICIENT indicates that if a user can be authenticated against Oracle Internet Directory, no further authentication is required. REQUIRED indicates that the authentication provider must authenticate the user even if another provider has already authenticated the user.

5. Set the order of the authentication providers and the values of their control flags as follows:

   a. **DefaultAuthenticator**: SUFFICIENT

   b. **OIDAuthenticator**: SUFFICIENT

   c. **OSSOIdentityAsserter**: SUFFICIENT

   d. **DefaultIdentityAssert**: This does not have a control flag.

6. If your application requires the user name to be in the same case as in Oracle Internet Directory (uppercase, lowercase, initial capitals), select **Use Retrieved User Name as Principal**.

7. Save the changes.

8. Activate the changes.

9. Restart Oracle WebLogic Server.

Enabling OSSO for Oracle SES

To enable OSSO for Oracle SES, perform the following steps:

1. In the Administration GUI, click **Configure Single Sign-On** link in the **Global Settings** tab.

2. Select the **OSSO** option, and click **Activate**.

3. Enter the following parameter values for OSSO:

   - **Hint cookie enabled**: Specify true to enable hint cookie. Specify false to disable it.

   - **Hint cookie name**: Specify a name for the hint cookie when hint cookie is enabled.

   - **Query invalid session return URL**: Specify the URL to display when the query application session expires.

   - **Query logout return URL**: Specify the URL to display after logging out of the query application.

4. Click **Finish**.

Configuring Secure Search with Oracle Access Manager Single Sign-On

You can implement an SSO authentication mechanism for Oracle SES by using Oracle Access Manager (OAM).

> **Note:** This section describes SSO configuration steps for OAM version **10g**. For OAM version **11g**, refer to *Oracle Secure Enterprise Search Release Notes*.

Ensure that the following components are installed:

- Oracle Access Manager 10.1.4.3.0 or later. See *Oracle Fusion Middleware Installation Guide for Oracle Identity and Access Management*.

- Oracle HTTP Server (OHS) 11g.

- Oracle Internet Directory (OID) 10.1.4.3.0 or later. See *Oracle Fusion Middleware Administrator's Guide for Oracle Internet Directory*. Also see "Adding OAM Identity Asserter" on page 11-30 for information on configuring OID.

- Oracle HTTP Server WebGate.

  You must install OAM and add an entry for WebGate in OAM before installing WebGate. *Oracle Fusion Middleware Installation Guide for Oracle Identity and Access Management* provides detailed information about installing WebGate. Follow the steps as provided in this guide. However, for some steps, such as creating a WebGate instance and installing WebGate, you must provide certain OAM SSO-specific parameters, as listed in "Installing and Configuring WebGate" on page 11-33.

To implement the OAM-SSO authentication on Oracle SES, you must configure OHS, Oracle SES, OID, and OAM.

To configure OAM SSO for Oracle SES:

1. Add OAM Identity Asserter as described in "Adding OAM Identity Asserter" on page 11-30.

2. Add OID Authenticator as described in "Adding OID Authenticator" on page 11-31.

3. Configure Oracle HTTP server as described in "Configuring Oracle HTTP Server" on page 11-32.

4. Install and Configure WebGate as described in "Installing and Configuring WebGate" on page 11-33.

5. Integrate OAM with Oracle SES as described in "Integrating OAM with Oracle SES" on page 11-34.

Adding OAM Identity Asserter

To add an OAM Identity Asserter to the domain, perform the following steps:

1. Log in to the WebLogic Server Administration Console.

2. Click **Security Realms**, *Default Realm Name*, **Providers**.

3. Click **New** under the Authentication Providers table.

4. Enter a name for the new provider, select its type, and then click **OK**. For example:

   Name: `OAM Identity Asserter`

   Type: `OAMIdentityAsserter`

5. Click the name of the newly added provider.

6. On the Common tab, set the appropriate values for common parameters and set the Control Flag to `SUFFICIENT`.

7. Save the changes.

8. Restart the WebLogic Server.

Adding OID Authenticator

To add an Oracle Internet Directory (OID) Authenticator to the domain, perform the following steps:

1. Log in to the WebLogic Server Administration Console.

2. Click **Security Realms**, *Default Realm Name*, **Providers**.

3. Click **New** under the Authentication Providers table.

4. Enter a name for the new provider, select its type, and then click **OK**. For example:

   Name: `OID Authenticator`

   Type: `OracleInternetDirectoryAuthenticator`

5. Save the changes.

6. Click the newly added authenticator to see the Settings page. Retain the default settings; do not change the Control Flag until you have verified that the Oracle Internet Directory configuration is valid.

7. On the Common tab, specify the following required settings:

   Propagate Cause For Login Exception: Check.

   Principal: LDAP administrative user. For example: cn=orcladmin

   Host: The Oracle Internet Directory host name

   Use Retrieved User Name as Principal: Check

   Credential: LDAP administrative user password. For example: password

   Confirm Credential: For example: password

   Group Base DN: Oracle Internet Directory group search base

   User Base DN: Oracle Internet Directory user search base.

   Port: Oracle Internet Directory port

8. Save the changes.

9. Restart the WebLogic Server.

After adding OAM Identity Asserter and OID Authenticator as authentication providers, perform the following steps:

1. Log in to the WebLogic Administration Console.

2. Click **Security Realms**, **Default Realm Name**, **Providers**.

3. Select the **Users and Groups** tab to see the list of users and groups contained in the configured authentication providers.

   You should see user names from the Oracle Internet Directory configuration, which implicitly verifies that the configuration is working.

4. If the Oracle Internet Directory instance is configured successfully, you can change the control flag.

If the Oracle Internet Directory authentication is sufficient for an application to identify the user, then choose the SUFFICIENT flag. SUFFICIENT indicates that if a user can be authenticated against Oracle Internet Directory, no further authentication is required. REQUIRED indicates that the authentication provider must authenticate the user even if another provider has already authenticated the user.

5. Set the order of the authentication providers and the values of their control flags as follows:

   a. **DefaultAuthenticator**: SUFFICIENT

   b. **OIDAuthenticator**: SUFFICIENT

   c. **OAMIdentityAsserter**: SUFFICIENT

   d. **DefaultIdentityAssert**: This does not have a control flag.

6. If your application requires the user name to be in the same case as in Oracle Internet Directory (uppercase, lowercase, initial capitals), select **Use Retrieved User Name as Principal**.

7. Save the changes.

8. Activate the changes.

9. Restart Oracle WebLogic Server.

Configuring Oracle HTTP Server

To configure OHS, perform the following tasks:

1. Edit mod_wl_ohs.conf to include the following. The file is available at *ORACLEOHS_HOME*/instances/*instance1*/config/OHS/ohs1/, where *instance1* refers to the instance name of OHS.

```
<IfModule weblogic_module>
     WebLogicHost [SES host name]
     WebLogicPort [SES HTTP port]
     WLLogFile Convenient Location of the log
</IfModule>

<Location /search/query>
     SetHandler weblogic-handler
</Location>

<Location /search/admin>
     SetHandler weblogic-handler
</Location>

# For monitor SES URL
<Location /monitor>
     SetHandler weblogic-handler
</Location>

# For Help links in Admin side
<Location /search/ohw>
     SetHandler weblogic-handler
</Location>
```

For example, if your SES host is sesHost and the port is 8001:

```
<IfModule weblogic_module>
```

```
        WebLogicHost sesHost
        WebLogicPort 8001
        WLLogFile /scratch/exampleuser/weblogic.log
</IfModule>

<Location /search/query>
        SetHandler weblogic-handler
</Location>

<Location /search/admin>
        SetHandler weblogic-handler
</Location>

<Location /monitor>
        SetHandler weblogic-handler
</Location>

<Location /search/ohw>
        SetHandler weblogic-handler
</Location>
```

2. Edit `httpd.conf` located at *ORACLEOHS_HOME*`/instances/instance1/config/OHS/ohs1/` to include the following at the end of the file:

   ```
   # Include configuration for mod_weblogic
   include "${ORACLE_INSTANCE}/config/${COMPONENT_TYPE}/${COMPONENT_NAME}/mod_wl_ohs.conf"
   ```

   Ensure that this line of code is on a single line.

3. Restart the HTTP server.

   ```
   $ORACLEOHS_HOME/instances/instance1/bin/opmnctl restartproc process-type=OHS
   ```

Installing and Configuring WebGate

A WebGate is a Web server plug-in that is shipped out-of-the-box with Oracle Access Manager (OAM). The WebGate intercepts HTTP requests from users for Web resources and forwards them to the Access Server for authentication and authorization. See *Oracle Fusion Middleware Installation Guide for Oracle Identity and Access Management* for more information on installing a WebGate.

While installing WebGate, you must configure some parameters for the OAM SSO authentication.

Creating a WebGate Instance

Provide the following values while defining a WebGate instance in the Access System Console:

- **AccessGateName**: Set as `SESAccessGate`

- **Description**: Set as `Secure Enterprise Search Access Gate`

- **HostName**: This is the host name on which OHS is installed.

- **AccessGate Password**: Set a password.

- **Port**: This is the port number set during OHS installation.

- **Transport Security**: Set to `Open`.

- **Preferred HTTP Host**: The domain for OHS. For example, if the OHS hostname is `myhost.oracle.com`, then the domain is `oracle.com`.

- Ensure that **Access Management Service** is on.

Installing WebGate

Provide the following parameters while specifying the WebGate configuration details:

- **WebGate ID**: Enter `SESAccessGate`.

- **WebGate Password**: The same as **AccessGate** password.

- **Access Server ID**: Obtain this from Access System Console.

- **DNS hostname**: Obtain this from Access System Console.

- **Port number**: Obtain this from Access System Console.

Updating the WebGate Web Server Configuration

Use the option to automatically update the Web Server configuration.

Integrating OAM with Oracle SES

Perform the following tasks:

1. Create a login page for Oracle HTTP Server. For example, *ORACLEOHS_HOME*/ohs/htdocs/login/login.html:

```
<html>
<head>
<title>SES-OAM Test Login Page</title>
<body bgcolor="white">
<h1 align="center">SES-OAM SSO Login Page: Sign-In</h1>
<form method="POST" action="/myaction/test.html">
  <table border="0" cellspacing="5">
    <tr>
      <th align="right">Username:</th>
      <td align="left"><input type="text" name="usernamevar"></td>
    </tr>
    <tr>
      <th align="right">Password:</th>
      <td align="left"><input type="password" name="passwordvar"></td>
    </tr>
    <tr>
      <td align="right"><input type="submit" value="Log In"></td>
      <td align="left"><input type="reset"></td>
    </tr>
  </table>
</form>
</html>
```

2. Define a form-based authentication in OAM Policy Manager:

   a. From `http://OAMHost:OAMPort/access/oblix`, select **Access System Console**, then **Access System Configuration**, and then **Authentication Management**.

   b. Create Form Login method with the following options:

   Name: OAMFormLogin

   Description: OAM Form-based login

   Level: 1

Challenge Method: Form

Challenge Parameter

`form`: /login/login.html

`creds`: usernamevar passwordvar

`action`: /myaction/test.html

`passthrough`: no

SSL Required: No

Enabled: Yes

c. Set up the following plugins under the **Plugins** tab:

credential_mapping:

```
obMappingBase="o=company,c=us",obMappingFilter="(&(&(objectclass=gensiteorg
person)(genuserid=%usernamevar%))(|(!(obuseraccountcontrol=*))(obuseraccoun
tcontrol=ACTIVATED)))"
```

validate_password:

```
obCredentialPassword="passwordvar"
```

where `obMappingBase` is the base DN in the user search in the LDAP directory
server, and `obMappingFilter` is the LDAP filter used to search for a user with
a given user ID. The directory login attribute is an attribute defined in the
Identity System using a Semantic login type.

d. Ensure that a default step exists in the **Steps** tab to use the `credential_`
`mapping` and `validate_password` plugins.

3. Create a policy in the Policy Manager to protect the query application login link
using the form authentication created in the previous step:

a. From `http://OAMHost:OAMPort/access/oblix`, select **Policy Manager**, and
then **Create Policy Domain**.

b. Protect an HTTP resource with `/search/query/formlogin.uix` as the URL
prefix.

c. In the **Authorization Rules** tab, add the role `myrole`. Also set the following:

Enabled: Yes

Allow takes precedence: Yes

d. Under **Actions** tab for `myrole`, first add the following return action:

Type: `HeaderVar`

Name: `OAM_REMOTE_USER`

Return Attribute: `uid`

e. Under **Allow Access** tab, ensure that anyone is allowed access.

f. Enable the new policy under **My Policy Domains**.

g. Click **Default Rules**, and under **Authentication Rule**, add a rule to use the
form login scheme as the **Authentication Scheme**.

h. Under **Authorization Expression**, ensure that `myrole` is selected for **Default
Rules**.

4. Create a policy in Policy Manager to protect the HTTP resource `/search/query` with the **Anonymous Authentication** option. The steps are identical to the previous step. However, for step 3g, the form login scheme must be **Anonymous Authentication** under **Authentication Rule**.

5. Turn on the OAM Single Sign-On using the Administration GUI by navigating to the **Global Settings > Configure Single Sign-On** page, selecting the **OAM** option, and clicking **Activate**.

6. Enter the following parameter values for OAM SSO:

   ■ **Admin logout return URL**: Specify the URL to display after logging out of the administration application.

   ■ **Query invalid session return URL**: Specify the URL to display when the query application session expires.

   ■ **Query logout return URL**: Specify the URL to display after logging out of the query application. Specify this URL format depending upon the WebGate version.

   Note: The format of the logout return URLs for administration and query applications must be as follows:

   For WebGate 10g: `/oamsso/logout.html?end_url=`*`return_url`*

       where,

       `return_url` is the page that is displayed after logging out of the application. For example, to return to the default query application page, set *`return_url`* to `/search/query`.

   For WebGate 11g: *`oam_logout_url`*`?end_url=`*`return_url`*

       where,

       `oam_logout_url` is the OAM Logout URL. For example, `http://`*`oam_server_host:oam_server_port`*`/oam/server/logout`

       `return_url` is the page that is displayed after logging out of the application. For example, to return to the default query application page, set *`return_url`* to `http://ses_host:ses_port/search/query`.

   ---

7. Click **Finish**.

> **Note:**
>
> - If you protect the Oracle SES instance using OAM SSO, then it is recommended to protect both the Oracle SES applications - Administration application and Query application.
>
> - If you chose to protect only one of the Oracle SES applications using OAM SSO, that is, either the Administration application and or the Query application, then you must configure the logout URL parameter for the *unprotected* application using either the Administration GUI or the Administration API.
>
>   For example, if the Administration application is *unprotected*, then the Administration URL must be set to be *excluded* from OAM resource, and the Administration Logout URL must be set to a non-SSO logout URL, such as `/search/admin/control/logout.jsp`.
>
> - If you protect the Administration application using OAM SSO, then you must either add the *searchsys* user to the OID used for SSO, or grant the `SearchAdminSuperUserRole` role to an existing user in OID. Also set the appropriate Administration Logout URL for SSO.

> **See Also:** "Configuring Centralized Logout for 11g [or 10g] WebGate with OAM 11g Servers" in the *Oracle Access Manager Access Administration Guide*

SSL and HTTPS Support in Oracle Secure Enterprise Search

For SSL support, Oracle SES uses JSSE, a highly-customizable SSL package included in Sun Microsystem's J2SE. Oracle SES uses SSL for many operations, some acting as the SSL client, and others acting as the SSL server.

Oracle SES can crawl HTTPS-based URLs, and the Oracle SES middle tier can be configured to support HTTPS-based access. HTTPS refers to HTTP running over a secure socket layer (SSL).

Understanding SSL

SSL is an encryption protocol for securely transmitting private content on the internet. Using SSL, two parties can establish a secure data channel. SSL uses a cryptographic system that uses two keys to encrypt data: a public key and a private key. Data encrypted with the public key can only be decrypted using the private key, and vice versa.

In SSL terms, the party that initiates the communication is considered the client. During the SSL handshake, authentication between the two parties occurs. The authentication can be one sided (server authentication only) or two sided (server and client authentication).

Server authentication is more common. It happens every time a Web browser accesses a URL that starts with HTTPS. Because of server authentication, the client can be certain of the server's identity and can trust that it is safe to submit secure data such as login username and password to the server.

The following list defines some common terms related to SSL:

- **Keystore**: A repository that includes the following:

  - Certificates identifying trusted entities. If a keystore contains certificates of only trusted entities, then it is referred to as a *truststore*.

  - Private-key and the matching certificate. This certificate is sent as a response to SSL authentication challenges.

- **Certificate**: A digital identification of an entity that contains the following:

  - SSL public key of the server

  - Information about the server

  - Expiration date

  - Digital signature by the issuer of the certificate used to verify the authenticity of the certificate

- **Certificate authority (CA)**: A well known and trusted entity (for example, VeriSign or Thawte). CAs are usually the issuers of other certificates.

- **Root certificate**: A self-signed certificate where the issuer is the entity that the certificate represents. CA certificates are typically root certificates.

- **Certificate chain**: This chain consists of the certificate, its issuer, the issuer of the issuer, and so on, all the way to the root certificate. If one certificate in the chain is trusted (that is, it is in the keystore), then the rest of the certificate can be verified for authenticity. This makes it possible for a keystore to contain only a few well-known and trusted root certificates from which most other certificates originate.

Every SSL connection starts with the SSL handshake. These are the basic steps:

1. The client contacts the server to establish a SSL connection.

2. The server looks in its keystore for its own SSL certificate and sends it back to the client.

3. The client checks its keystore to see if it trusts the server or any of the entities in the server's certificate chain. If not, then the handshake is aborted. Otherwise, the client positively identifies the server and deems it trusted. The expiration date of the certificate is also checked, and the name on the certificate is matched against the domain name of the server.

4. If the server is configured to require client authentication, then the server asks the client to identify itself, so the mirror image of steps 2 and 3 takes place.

5. Session keys are generated and used for encrypting the transmitted data.

Oracle strongly recommends that you use an SSL-protected channel to transmit password and other secure data over networks.

Typically, the following components transmit password and other secure data over a network:

- Federation

- Connectors

- Authorization Plugin

- Identity plugin

- Suggested content

- Web Service APIs

Managing the Keystore

The keystore is populated with the root certificates representing well known certificate authorities. Most SSL-enabled Web sites use certificates that originate or chain from these main root certificates. See "Oracle SES Acting as an SSL Server" on page 40 for more information about managing the keystore.

> **See Also:** The Java Secure Socket Extension (JSSE) Reference Guide at
>
> http://docs.oracle.com/javase/6/docs/technotes/guides/security/jsse/JSSERefGuide.html

Importing SSL Certificates into the Java Virtual Machine

For connectors that interact with external SSL-enabled repositories at crawl time and query time, you must import the SSL certificate into the keystore location of the Oracle SES middle tier, that is, into the WebLogic Server directory *wls_home*/server/lib.

Maintaining a Keystore

Depending on the requirement, the keystore might need maintenance. For example:

- If a main root certificate has expired, then it must be replaced by a new issue.

- If Oracle SES must trust another SSL-enabled peer whose certificate does not originate from a root certificate, then the peer's certificate, or one from its chain, must be added to the keystore.

- To enable SSL in the Oracle SES middle tier, Oracle SES must act as an SSL server, and that calls for the keystore to contain a private key and the corresponding certificate with the public key. (The same holds true for the SSL client role where the server requires client side SSL authentication.)

Maintenance of the keystore can be done using Sun Microsystem's keytool program, which ships with J2SE. You can find this utility under *java_home*/bin. Third-party keytool GUI wrapper programs are available.

> **See Also:** For detailed instructions on how to add, remove, or update certificates, generate keys, and create new keystores with a keytool:
>
> http://docs.oracle.com/javase/6/docs/technotes/tools/solaris/keytool.html

Oracle SES Acting as an SSL Client

Oracle SES acts as the SSL client in the following situations:

- The crawler accesses a data repository that uses SSL (for example, HTTPS Web sites).

- The form registration wizard in the Oracle SES Administration GUI accesses HTTPS URLs.

- Oracle SES federates queries to other SSL-enabled search services (for example, an SSL-enabled Oracle SES instance).

  Note that for Oracle SES 11.1.2 instances, the broker and the endpoint do not have to exchange any certificates when the broker tries to create a federated source using the HTTPS Web service URL. This is because the Oracle SES instances share the same default certificates in the trusted store.

If you crawl an SSL-enabled Web site whose SSL key is not in the SSL keystore, the following error occurs:

```
@ javax.net.ssl.SSLHandshakeException:
sun.security.validator.ValidatorException: No trusted certificate found
```

To fix this error, you can add the key to the Oracle SES keystore.

To add an SSL certificate to the Oracle SES keystore:

1. Access the page in a browser, and accept the SSL certificate when prompted.

2. View the certificate through your browser options.

3. Import the certificate into the Oracle SES keystore.

4. Try the crawl again.

The following sections explain how to import certificates.

Oracle SES Acting as an SSL Server

Oracle SES acts as the SSL server when the Oracle SES middle tier, configured to use SSL, responds to HTTPS requests. The Oracle SES crawler connects to SSL-enabled sites using the JSSE package, which contains a keystore with a few default certificates from well known CAs.

This section contains the following topics:

- Configuring Oracle Secure Enterprise Search to Use SSL
- Configuring Oracle HTTP Server to Use SSL

Configuring Oracle Secure Enterprise Search to Use SSL

When Oracle SES is used along with the Oracle HTTP Server, Oracle recommends that Oracle SES be configured to use SSL with client-side authentication for communication with the Oracle HTTP Server. Furthermore, use a keystore other than the default one. Oracle recommends that you create separate identity and trust keystores.

The communication channel between the client and Oracle SES is by default not SSL-enabled and not encrypted.

Perform the following steps:

1. Create a new keystore. This step is optional but Oracle recommends it. To create the keystore:

    a. Log in to the WebLogic Server Administration Console.

    b. Expand the **Environment** button and click **Servers**. This takes you to the configuration page for the servers.

    c. Click the name of the server for which you want to configure SSL.

    d. Click the **keystores** tab.

    e. From the keystores list, select **Custom Identity and Custom Trust**.

    f. In the **custom identity keystore** field, add the complete path and name of the new keystore. To create a new keystore SESIdentity.jks, add the path and name *wls_home*/server/lib/SESIdentity.jks to the keystore field.

    g. Set the custom identity keystore type to be jks. Set a pass phrase for the store.

    h. In the custom trust field, add the complete path and name of the new keystore. To create a new keystore `SESTrust.jks`, add the path and name `wls_home`/server/lib/SESTrust.jks to the keystore field.

    i. Set the custom identity keystore type to be `jks`. Set a pass phrase for the store.

    j. Click **Save** to create the new keystores.

2. Create new certificates for the identity and trust keystores. Use Sun Microsystem's keytool utility to perform the following steps:

    a. Generate the key for the identity keystore by executing the following `keytool` command from the *java_home*/bin directory:

```
>keytool -genkey -alias [MyCertificateAlias] -keyalg RSA -keysize 1024
-dname ["My DN"] -keypass [MyKeyPass ] -keystore [MyKeyStore ]  -storepass
[PasswordOfTheKeystoreCreatedAbove] -storetype [StoreTypeCreatedAbove]
```

For example,

```
> keytool -genkey -alias sescert -keyalg RSA -keysize 1024 -dname
"CN=example0123.us.mycompany.com,OU=ses,O=oracle,C=us" -keypass welcome1
-keystore wls_home/server/lib/SESIdentity.jks -storepass welcome1
-storetype jks
```

The above command creates a certificate with the alias `sescert` and the given dn and keypass `welcome1`. It uses `SESIdentity.jks` as the keystore, which is the same as the one created in step 1. The storepass and the storetype are the same as supplied in step 1.

    b. Generate the key for the trust keystore by executing the following `keytool` command from the *java_home*/bin directory:

```
> keytool -genkey -keyalg RSA -alias sescert -keysize 1024 -dname
"CN=example0123.us.mycompany.com,OU=ses,O=oracle,C=us" -keypass welcome1
-keystore wls_home/server/lib/SESTrust.jks -storepass welcome1 -storetype
jks
```

    c. Certify the generated keys by running the following command:

```
> keytool -selfcert -alias sescert -keyalg RSA -validity 2000 -keypass
welcome1 -keystore wls_home/server/lib/SESIdentity.jks -storepass welcome1
```

The command uses the alias, keypass, and the keystore location supplied in step 2.a. The store pass is the password of the store.

To self certify the keystore, run the following command:

```
> keytool -selfcert -alias sescert -keyalg RSA -validity 2000 -keypass
welcome1 -keystore wls_home/server/lib/SESTrust.jks -storepass welcome1
```

Note: In addition to using Sun Microsystem's Keytool utility to self-sign the generated key, you can use any of the options mentioned here: http://docs.oracle.com/cd/E12840_01/wls/docs103/secmanage/identity_trust.html

---

3. Configure Oracle SES to use the generated key:

    a. Log in to the WebLogic Server Administration Console.

    b. Select the server for which you want to configure SSL by expanding the **Environment** button and clicking on **Servers**. The configuration page for servers is displayed.

    c. Click the **ssl** tab.

    d. The private key location is set to **from Custom Identity Keystore**.

    e. In the **Private Key Alias** field, provide the private key alias. This is the alias specified in step 2a.

    f. Provide the private key pass phrase that you specified in step 2a.

    g. Save the changes.

4. (Optional) Enable two-way-SSL for Oracle SES:

    a. Log in to the WebLogic Server Administration Console.

    b. Select the server for which you want to configure SSL by expanding the **Environment** button and clicking on **Servers**. The configuration page for servers is displayed.

    c. Click the **SSL** tab.

    d. Click **Advanced** link present at the bottom of the page.

    e. Set the parameter **Two Way Client Cert Behavior** value to **Client Certs Requested and Enforced**.

    f. Save the changes and restart the SSL channel.

5. (Optional) Enable WebLogic plugin when the WebLogic Server is used along with the Oracle HTTP Server that uses WebLogic proxy plugin:

    a. Log in to the WebLogic Server Administration Console.

    b. Select the server for which you want to configure SSL by expanding the **Environment** button and clicking on **Servers**. The configuration page for servers is displayed.

    c. Click the **General** tab.

    d. Click **Advanced** link present at the bottom of the page.

    e. Select **Weblogic plugin enabled**.

    f. Save the changes and restart the Oracle SES server.

6. Enable SSL for Oracle SES:

    a. Log in to the WebLogic Server Administration Console.

    b. Select the server for which you want to configure SSL by expanding the **Environment** button and clicking on **Servers**. The configuration page for servers is displayed.

    c. Click the **General** tab.

    d. Select **SSL Listen Port Enabled** and provide a port number. The default port is `7002`.

    e. Save the changes.

    f. Click the **Control** tab. You can access the control tab by expanding the **Environment** button and clicking on **Servers**.

    g. From the Control tab, restart SSL.

Configuring Oracle HTTP Server to Use SSL

Configuring Oracle HTTP Server (OHS) to use SSL is a multistep process involving configuration of the server, modification of certain `.conf` files, and exchange of certificates.

To configure Oracle HTTP Server to require SSL:

1. Configure the Oracle HTTP server:

   a. From `ORACLEOHS_HOME`/bin, run `owm`, where, `ORACLEOHS_HOME` is the root directory of the Oracle HTTP Server (OHS).
   This opens Oracle Wallet Manager, which is used to create the certificate for OHS.

   b. Click **Wallet** and then click **New**.

      If you get a message indicating that the default directory is not set, click **Continue**.

   c. Provide a password for the wallet. Click **No** for the option to configure user certificate request.

   d. Click **Wallet** and then click **Save As**. Save the wallet to the directory `ORACLEOHS_HOME`/instances/*instanceName*/config/OHS/*componentName*/keystores/myWallet. This creates a new wallet with the name `myWallet` for the Oracle HTTP server.

      `instanceName` and `componentName` are specified during the installation of Oracle HTTP Server.

   e. Create a key-cert pair (a user certificate) using the following command from `ORACLEOHS_HOME`/bin

      ```
      > orapki wallet add -wallet [walletPath] -dn ["myDN"] -keysize 1024 -self_
      signed -validity 720
      ```

      For example,

      ```
      > orapki wallet add -wallet $ORACLEOHS_
      HOME/instances/instance1/config/OHS/ohs1/keystores/myWallet -dn
      CN=example0123.us.mycompany.com,OU=ohs.ses,O=oracle,ST=ca,C=US -keysize
      1024 -self_signed -validity 720
      ```

      The command adds a user certificate with the given dn and the wallet located at `ORACLEOHS_HOME`/instances/instance1/config/OHS/ohs1/keystores/myWallet. Note that `instance1` is the name of the instance provided during installation and `ohs1` is the name of the component provided during installation.

   f. Go back to the OWM utility and reopen the wallet. To reopen it, close and open the wallet by selecting the correct directory. You should now see `Certificate:[Ready]` under the wallet.

   g. Save the changes.

   h. Double-click **Certificate:[Ready]**, click the **Operations** tab, and select **export user certificate**. Export the user certificate file (`/tmp/OHSIdentityCertificate.crt`) to a suitable location.

2. Edit the file `ssl.conf` located at `ORACLEOHS_HOME`/instances/*instanceName*/config/OHS/*componentName*/.to include the following. Note that *instanceName* and *componentName* are specified during the installation of Oracle HTTP Server.

```
<VirtualHost*:dddd>
<IfModule mod_weblogic.c>
   WebLogicHost [SESHost]
   WebLogicPort [SESPort]
   Debug ALL
   WLLogFile [Location of the log file]
   SecureProxy On
   WlSSLWallet "MyWalletLocation"
<Location /weblogic>
   SetHandler weblogic-handler
   PathTrim /weblogic
</Location>
<Location /console>
   SetHandler weblogic-handler
</Location>
</IfModule>
</VirtualHost >
```

For example, if the host is `sesHost`, the port is `7002`, and the wallet is located at `Oracle_Instance`/config/`Component_Type`/`Component_Name`/keystores/myWallet, then the following configuration file is helpful:

```
<IfModule mod_weblogic.c>
   WebLogicHost sesHost
   WebLogicPort 7002
   Debug ALL
   WLLogFile /scratch/exampleuser/Certificates/weblogic.log
   SecureProxy On
   WlSSLWallet "${ORACLE_INSTANCE}/config/${COMPONENT_TYPE}/${COMPONENT_
NAME}/keystores/myWallet"
<Location /weblogic>
   SetHandler weblogic-handler
   PathTrim /weblogic
</Location>
<Location /console>
   SetHandler weblogic-handler
</Location>
</IfModule>
```

3. Edit the file `mod_wl_ohs.conf` located at `ORACLEOHS_HOME`/instances/`instanceName`/config/OHS/`componentName`/ to include the following:

```
<IfModule weblogic_module>
WebLogicHost [SES host name]
  WebLogicPort [SES HTTP port]
  Debug ON
  WLLogFile [Location of the log]
</IfModule>
<Location /search/query>
  SetHandler weblogic-handler
</Location>
<Location /search/admin>
  SetHandler weblogic-handler
</Location>
# For monitor SES URL
<Location /monitor>
SetHandler weblogic-handler
</Location>
# For Help links in Admin side
<Location /search/ohw>
```

```
SetHandler weblogic-handler
</Location>
```

For example if the Oracle SES host is `sesHost` and the port is `8001`, then the file would contain:

```
<IfModule weblogic_module>
  WebLogicHost sesHost
  WebLogicPort 8001
  Debug ON
  WLLogFile /scratch/exampleuser/weblogic.log
</IfModule>
<Location /search/query>
  SetHandler weblogic-handler
</Location>
<Location /search/admin>
  SetHandler weblogic-handler
</Location>
<Location /monitor>
SetHandler weblogic-handler
</Location>
<Location /search/ohw>
SetHandler weblogic-handler
</Location>
```

4. Exchange the certificates for OHS and Oracle SES WebLogic Server. To exchange them, use Oracle Wallet Manager to import and export certificates from and to the wallet, and use the Java keytool for the Oracle SES keystore. While importing a certificate, ensure that it is self-signed. If not, then you must import any of the certificates in the chain. See "Understanding SSL" on page 11-37 for more information about certificate chains.

Perform the following steps to exchange certificates:

a. Export the SESIdentity key generated in step 2a of Configuring Oracle Secure Enterprise Search to Use SSL to a suitable location by running the following command from the Oracle SES directory *java_home/bin*:

```
> keytool -export -alias sescert -keystore wls_
home/server/lib/SESIdentity.jks -file /tmp/SESIdentityCertificate.crt
```

The above command exports the certificate with the alias `sescert` and the keystore created in step 2a of Configuring Oracle Secure Enterprise Search to Use SSL to the file `/tmp/SESIdentityCertificate.crt`.

b. Import the exported OHS certificate created in step 1h to Oracle SES. To import it, run the following command from the Oracle SES directory *java_home/bin*:

```
> keytool -file [LocationOfOHSIdentityCertificate]  -alias [MyOHSCerAlas]
-import -trustcacerts -keystore [LocationofSESTrustStore] -storepass
[MyPasswordForTheTrustStore] -storetype jks
```

For example, if the location of the exported OHS identity certificate is `tmp/OHSIdentityCertificate.crt`, the Oracle SES trust store is at *wls_home*/server/lib/SESTrust.jks, the store password is `welcome1`, and the alias is `ohsCert`, then run the following:

```
> keytool -file tmp/OHSIdentityCertificate.crt  -alias ohsCert -import
-trustcacerts -keystore wls_home/server/lib/SESTrust.jks -storepass
welcome1 -storetype jks
```

    c. Import the Oracle SES certificate into OHS wallet. The Oracle SES certificate is the file `SESIdentityCertificate.crt` exported in step 4a. To import this certificate, from the Oracle Wallet Manager utility, click **Operations** and select **Import Trusted Certificate**. Navigate to the location of the exported Oracle SES certificate (`/tmp/SESIdentityCertificate.crt`), and import it as a trusted certificate.

    d. Restart Oracle HTTP Server. Before restarting the server, ensure that the **Auto Login** option is enabled in Oracle Wallet Manager. The restart fails if the option is not enabled.

      To restart the server, run the following command from `ORACLEOHS_ HOME/instances/instance1/bin/`:

```
opmnctl restartproc process-type=OHS
```

    e. Restart SSL for the WebLogic Server by using the control page of the server.

      To access the control page, click **Environment**, then **Server**, and then **Control**.

> **See Also:** *Oracle Database Advanced Security Administrator's Guide* for more information about Oracle Wallet Manager
>
> *Oracle Database Advanced Security Administrator's Guide* for more information about the orapki utility
>
> *Oracle HTTP Server Administering a Standalone Deployment Based on Apache 2.0* for more information about enabling SSL for OHS

Using Windows Native Authentication

To log in to the Oracle SES application on the Windows platform, you can choose to implement the user authentication mechanism at the Oracle SES application layer, which involves logging in through the Oracle SES login page, or at the Windows operating system layer.

If you enable authentication at the Windows OS layer, then when you log in to a computer using your Windows user credentials, you can automatically access the Oracle SES application as a logged user. You are not required to provide additional authentication credentials to access the Oracle SES application. This is implemented using the Single Sign-On authentication mechanism. To implement this mechanism, the Oracle SES Identity plug-in must be integrated with the Active Directory that manages Windows user authentication. The authentication is implemented using the Kerberos encryption mechanism.

> **Note:** In the earlier Oracle SES release (11.1.2.2), when Oracle SES is integrated with Windows Native Authentication (WNA), the user is logged into the query application automatically regardless of the query application secure mode. This behavior is changed in Oracle SES 11.2.2.2. In Oracle SES 11.2.2.2:
>
> - If the secure mode of query application is "require login for both public and secure content", then a user is automatically logged into the query application. After logging out, the user is redirected to the login page.
>
> - If the secure mode of query application is "require login for secure content", then a user must again provide authentication credentials to login to the query application. After logging out, the user is redirected to the default search page.

Activating Windows Native Authentication

To activate Windows authentication, you must perform the following steps:

1. Configure the Active Directory. See "Configuring the Active Directory" on page 11-47.

2. Activate the Active Directory plug-in in Oracle SES. See "Activating the Active Directory Identity Plug-in" on page 11-10.

3. Enable WNA SSO for Oracle SES as described in "Enabling WNA SSO for Oracle SES" on page 11-52.

Configuring the Active Directory

The Active Directory is available on the Windows Server. As a first step, configure this active directory.

To configure the Active Directory:

1. In the Active Directory server, create a user account called `seswna` for the Oracle SES instance:

   a. Select **New**, and then **User**.

   b. Specify a password for the user. Do *not* select the **User must change password at next logon** option.

2. Configure the new user account to comply with Kerberos protocol. Ensure that the user account's encryption type is set to `DES` and the user account requires Kerberos pre-authentication. To implement this:

   a. Right-click the user name and select **Properties**.

   b. Click the **Account** tab and select **Use DES encryption types for this account**.

   c. Reset the user password. Right-click the user name, select **Reset Password**, and reenter the same password that you set earlier. (This step is recommended because setting the encryption type may corrupt the password that you set initially).

3. Create the Service Principal Names (SPNs) for the user account by using the `setspn` utility. Enter the following command:

   ```
   setspn -a HTTP/<ses-host-name> seswna
   ```

where <ses-host-name> must be a fully qualified network address like
`sesmachine.us.xyz.com`.

4. Create a user mapping using the `ktpass` utility. Enter the following command:

```
ktpass -princ HTTP/<ses- host-name>@<ad-domain-name> -pass <mapuser_password>
-mapuser seswna -out c:\temp\seswna.HTTP.keytab -crypto DES-CBC-CRC
```

5. Copy the keytab file `seswna.HTTP.keytab` to an Oracle SES middle tier directory. If
your Oracle SES environment has multiple middle tier host systems, then you
need to repeat the above two steps (3 and 4) for each middle tier host system, and
copy the `seswna.HTTP.keytab` file to each middle tier host system.

After configuring the `seswna` user in the Active Directory, you must set up an identity
plug-in for Active Directory. This identity plug-in must be configured to the Active
Directory where `seswna` is created. See "Activating the Active Directory Identity
Plug-in" on page 11-10 for more information about activating the Active Directory
plug-in.

Creating and Configuring JAAS Login File

To create and configure JAAS login file, perform the following steps:

1. Create a text file with the name `krb5Login.conf`.

2. Add the following text to the `krb5Login.conf` file:

```
com.sun.security.jgss.initiate {
com.sun.security.auth.module.Krb5LoginModule required
principal="HTTP/ses_hostname@REALM.NAME" useKeyTab="true"
keyTab="keytab_file_name" storeKey="true" debug="true";
};
com.sun.security.jgss.accept {
com.sun.security.auth.module.Krb5LoginModule Required
principal="HTTP/ses_hostname@REALM.NAME" useKeyTab="true"
keyTab="keytab_file_name" storeKey="true" debug="true";
};
```

where, *keytab_file_name* must be replaced with the file path of
`seswna.HTTP.keytab` file.

3. Copy the `krb5Login.conf` file to an Oracle SES middle tier directory. If your
Oracle SES environment has multiple middle tier host systems, then copy the
`krb5Login.conf` file to each middle tier host system.

Adding Negotiate Identity Assertion Provider

To add Negotiate Identity Assertion provider, perform the following steps:

1. Log in to the WebLogic Server Administration Console.

2. Acquire lock to make changes.

3. Expand the **Security > Realms** nodes.

4. Select the name of the realm.

5. Expand the **Providers > Authentication Providers** nodes.

6. On the **Authenticators** tab, click **Configure a new Negotiate Pass Identity
Asserter**.

7. Specify the Negotiate Identity Assertion provider name, and select **Negotiate
Identity Asserter** type as the provider type.

8. Save the changes.

Adding Active Directory Authenticator

To add Active Directory Authenticator, perform the following steps:

1. Log in to the WebLogic Server Administration Console.

2. Acquire lock to make changes.

3. Expand the **Security > Realms** nodes.

4. Select the name of the realm.

5. Expand the **Providers > Authentication Providers** nodes.

6. On the **Authenticators** tab, click **Configure a new ActiveDirectoryAuthenticator**.

7. Specify the Active Directory Authenticator name, and in the Common Authentication Provider Settings page, change the value for **Control Flag** from REQUIRED to SUFFICIENT.

8. Save the changes.

9. Click the **Provider Specific** tab and edit the Active Directory LDAP settings:

- **Connection**

  - Host: The Windows system name on which the Active Directory is installed.

  - Port: The port to which the Active Directory is listening to.

  - Principal: The LDAP DN for the user to connect to the Active Directory, for example, cn=jsmith,cn=users,dc=us,dc=xyzcorp,dc=com.

  - Credential/Confirm Credential: Password for the Principal.

- **Groups**

  - Group Base DN: The LDAP query to find groups in the Active Directory.

- **Users**

  - User Base DN: The LDAP query to find users in the Active Directory.

  - User Object class: User object class, for example, user.

  - Use Retrieved User Name as Principal: Select this check box.

  - User Attributes: Following are the user attributes containing the user related information in the Active Directory. You can either use the default settings or specify new settings for these attributes.

Attribute Name	Attribute Description	Default Setting	Specify New Setting
UserNameAttribute	User name in Active Directory.	cn	**<UserAttribute>**
AllUsersFilter	The LDAP search filter for all the users.	(&(cn=*)(objectclass=user))	(&(**<UserAttribute>** =*)(objectclass=person))
UserFromNameFilter	The LDAP search filter for a particular user.	(&(cn=%u)(objectclass=user))	(&(**<UserAttribute>**=%u)(objectclass=person))

- **General**

        – GUID attribute: The attribute used to define object GUIDs in Active Directory, for example, `objectguid`.

10. Save the changes.

11. Set the control flag of DefaultAuthenticator to `SUFFICIENT`.

12. Reorder the Authentication providers as follows:

    1. DefaultAuthenticator

    2. DefaultIdentityAsserter

    3. ActiveDirectoryAuthenticator

    4. NegotiateIdentityAsserter

13. Restart the WebLogic Server.

Configuring Startup Arguments for Kerberos Authentication

To configure startup arguments for Kerberos authentication, perform the appropriate steps depending upon whether the WebLogic Server is running *without* nodemanager or *with* nodemanager.

To configure startup arguments for Kerberos authentication with WebLogic Server running *without* a nodemanager:

1. Open the `setDomainEnv.sh` file in a text editor.

2. Add the following lines before the line beginning with `JAVA_OPTIONS="${JAVA_OPTIONS}"`:

```
WLS_WNA_JVM_OPTIONS="-Djava.security.auth.login.config=krb5Login.conf_file_
location -Djava.security.krb5.realm=Active_Directory_Realm
-Djava.security.krb5.kdc=KDC_hostname -Dweblogic.security.enableNegotiate=true"
```

where,

`conf_file_location` is the directory path of the JAAS login file `krb5Login.conf` created earlier as described in section "Creating and Configuring JAAS Login File" on page 11-48.

`Active_Directory_Realm` is the Active Directory realm name.

`KDC_hostname` is the Active Directory host name.

3. Update the line `JAVA_OPTIONS="${JAVA_OPTIONS}"` to:

```
JAVA_OPTIONS="${JAVA_OPTIONS} ${WLS_WNA_JVM_OPTIONS}"
```

4. Export the `WLS_WNA_JVM_OPTIONS` variable:

```
export WLS_WNA_JVM_OPTIONS
```

5. Restart the WebLogic Server.

If the WebLogic Server is running *with* nodemanager, then to configure startup arguments for Kerberos authentication, you can either execute the above described steps, or you can add the `WLS_WNA_JVM_OPTIONS` property value to the **Server Start** arguments for the Oracle SES server (`search_server1`).

Configuring Microsoft Internet Explorer to support WNA

To use WNA on Microsoft Internet Explorer, you must perform the following steps:

■ Configure Local Intranet Domains

- Configure Intranet Authentication
- Verify the Proxy Settings
- Additional Steps for Internet Explorer 6.0

Configure Local Intranet Domains To configure the local intranet domain, perform the following steps:

1. In Internet Explorer, select **Tools**, and then **Internet Options**.

2. From the Internet Options dialog box, select the **Security** tab.

3. Select **Local intranet** and then click **Sites**.

4. In the Local intranet dialog box, ensure that the options **Include all sites that bypass the proxy server** and **Include all local (intranet) sites not listed in other zones** are selected.

5. Click **Advanced**.

6. Add all relative domain names that are used for the WebLogic Server instances participating in the Single Sign-On configuration (for example, `myhost.example.com`) and click **Close**.

7. On the Local intranet dialog box, Click **OK**.

Configure Intranet Authentication To configure Intranet authentication, perform the following steps:

1. In Internet Explorer, select **Tools**, and then **Internet Options**.

2. From the Internet Options dialog box, select the **Security** tab.

3. Select **Local intranet**, and then click **Custom level**. This opens the Security Settings-Local Intranet Zone dialog box.

4. Under **User Authentication**, select **Automatic Logon only in Intranet Zone**. Note that this option prevents users from having to reenter log in credentials.

5. Click **OK**.

Verify the Proxy Settings If you have a proxy server enabled, then you must verify the proxy settings. To do this:

1. In Internet Explorer, select **Tools**, and then **Internet Options**.

2. From the Internet Options dialog box, select the **Connections** tab.

3. Click **LAN Settings** to open the Local Area Network (LAN) Settings dialog box.

4. Verify that the proxy server address and the port number are correct.

5. Click **Advanced** to open the Proxy Settings dialog box.

6. Ensure that the required domain names are entered in the **Exceptions** field.

Additional Steps for Internet Explorer 6.0 Additionally, for Internet Explorer 6.0, you must perform the following:

1. In Internet Explorer, select **Tools**, and then **Internet Options**.

2. From the Internet Options dialog box, select the **Advanced** tab.

3. Under **Security**, ensure that the option **Enable Integrated Windows Authentication** is selected.

4. If this option was not previously set, then restart the computer after setting the option.

Configuring Mozilla Firefox to Support WNA

In Mozilla Firefox, perform the following steps to use WNA:

1. In the Location bar, enter the string `about:config`. This opens the about:config page in Firefox.

2. In the Filter field, enter the string `network.negotiate`.

3. Set the preferences given in Table 11–7. To set the value for a preference, double-click the preference, and enter the value.

Table 11–7   WNA Configuration Preference for Firefox

Preference Name	Status	Type	Value
network.negotiate-auth.allow-proxies	default	boolean	true
network.negotiate-auth.delegation-uris	User set	string	http://,https://
network.negotiate-auth.gsslib	Default	string	<blank>
network.negotiate-auth.trusted-uris	User set	string	http://,https://
network.negotiate-auth.using-native-gsslib	Default	boolean	true

Enabling WNA SSO for Oracle SES

To enable WNA SSO for Oracle SES, perform the following steps:

1. In the Administration GUI, click **Configure Single Sign-On** link in the **Global Settings** tab.

2. Select the **WNA** option, and click **Activate**.

Changing the Master Encryption Key

A master encryption key is used to encrypt secure fields in Oracle SES. You can change this key if its security is compromised or for any other reason.

To change the master encryption key:

1. Stop all crawler schedules.

2. Close all middle-tier applications, except for the Monitor application.

3. Open an interactive session on the Oracle SES middle-tier computer.

4. Issue a `searchctl rollover_key` command. See the following description.

5. Restart the crawler and the middle-tier applications.

searchctl rollover_key

This command has the following syntax:

`searchctl rollover_key` *options*

Options have the format *keyword=value*:

ses_db_conn_str
Local JDBC connection string for the Oracle SES database. For example, `localhost:5555:ses1`. Required.

ses_admin_passwd
Oracle SES administrator's password. If you omit this password from the command, then you are prompted for it.

wls_admin_server
URL to the WebLogic Server Administration Console. For example, `t3://wls_example:8000`. Required.

wls_admin_user
User name of the WebLogic administrative user. (Required)

wls_admin_passwd
Password of the WebLogic administrative user. If you omit this password from the command, then you are prompted for it.

master_key
New master key. If you omit this option, a random master key is set.

The following command changes the master key to "testing123";

```
searchctl rollover_key ses_db_conn_str=localhost:5555:ses1 ses_admin_
passwd=password wls_admin_user=weblogic wls_admin_passwd=password wls_admin_
server=t3://asHost:8000 master_key=testing123
```

12

Administering Oracle SES Instances

This chapter provides information about tuning and general management of Oracle SES instances. It contains the following topics:

- Performing Backup and Recovery
- Tuning Crawl Performance
- Tuning Search Performance
- Using Command Line Tools
- Configuring Oracle Diagnostic Logging (ODL) for Oracle SES Server
- Viewing Oracle SES Server Log Files
- Monitoring Oracle SES Instance
- Integrating Oracle SES with Google Desktop
- Accessing Oracle WebLogic Server Administration Console

Performing Backup and Recovery

Oracle SES provides multiple methods for backup and recovery of the Oracle SES data in the event of a hardware failure. The backup and recovery process depends upon the mode of the Oracle SES installation.

Backup and Recovery for Oracle SES Installed with Database and Middle Tier

For the Oracle SES installed along with the Database and the middle tier, the backup and recovery can be done using the standard Oracle Recovery Manager (RMAN) utility. Refer to the *Oracle Database Backup and Recovery User's Guide* for more information.

Backup and Recovery for Oracle SES Installed on Existing Database or Middle Tier

For the Oracle SES installed on an existing Database or middle tier, the backup and recovery must be done according to the standard backup and recovery procedure for the Oracle Fusion Middleware. Refer to the *Oracle Fusion Middleware Administrator's Guide* for more information.

Post-recovery Steps for Oracle SES (for All Installation Modes)

The following Oracle SES specific additional steps must be executed after performing the recovery process explained in the earlier sections.

Update the WebLogic Server Connection Pool used by Oracle SES

You must update the WebLogic Server connection pool configuration, only if the database connection string used in the new database (used after recovery) is different from that of the old database (used before recovery).

To update the database connection string in the WebLogic Server connection pool:

1. Log in to the WebLogic Server Administration Console.

2. Under the **Change Center** in the left panel, click **Lock & Edit**.

3. Change the URL for SearchAdminDS data source:

   a. Under the **Domain Structure** in the left panel, select **search_domain > Services > Data Sources**. The Configuration tab for the Settings for SearchAdminDS is displayed in the main panel.

   b. In the Name column of the Data Sources table, click **SearchAdminDS**. The General tab of the Settings for SearchAdminDS is displayed.

   c. Select the Connection Pool tab.

   d. Enter the new database connection string in the URL field.

   e. Save the changes.

4. Repeat the above steps for the data sources - **SearchQueryDS**, **EssInternalDS**, **EssDS**, **EssXADS**, **mds-ESS_MDS_DS**, and **mds-owsm**.

5. Click **Activate Changes** under the **Change Center** in the left panel.

Update the Database Connection String in the Credential Storage Framework (CSF) used by the Crawler

You must update the database connection string in the Credential Storage Framework (CSF) used by the crawler, only if the database connection string used in the new database (used after recovery) is different from that of the old database (used before recovery).

To update the database connection string in the CSF:

1. Connect to the system where WebLogic Server middle tier is installed.

2. Go to *ses_home*/common/bin.

3. Start the WebLogic Server Administration Scripting Tool by executing the command `wlst.sh`.

4. Run the `connect()` command at the `wls/offline>` prompt. You will be prompted to enter the following values:

   - WebLogic Server user name

   - WebLogic Server password

   - WebLogic Server URL

5. Run the following command at the `wls:domain_name/serverConfig>` prompt:

   ```
   updateCred(map="oracle.search",key="SEARCH_DATABASE",user="new_database_
   connection_string",password="search")
   ```
   where, *new_database_connection_string* is the connection string URL for the new database.

> **Note:** The password value is not used in the above command. Use `search` as the password value.

Update the SEARCHSYS Schema Password in the Credential Storage Framework (CSF) used by the Crawler

You must update the SEARCHSYS schema password in the Credential Storage Framework (CSF) used by the crawler, only if the database schema password used in the new database (used after recovery) is different from that of the old database (used before recovery).

To update the SEARCHSYS schema password in the CSF:

1. Connect to the system where WebLogic Server middle tier is installed.

2. Go to *ses_home*/common/bin.

3. Start the WebLogic Server Administration Scripting Tool by executing the command `wlst.sh`.

4. Run the `connect()` command at the `wls/offline>` prompt. You will be prompted to enter the following values:

   - WebLogic Server user name

   - WebLogic Server password

   - WebLogic Server URL

5. Run the following command at the `wls:domain_name/serverConfig>` prompt:

   ```
   updateCred(map="oracle.apps.security",key="FUSION_APPS_ECSF_SES_ADMIN-KEY",
   user="searchsys",password="password")
   ```

   where,

   password is the new password for SEARCHSYS schema, and the Oracle SES credentials for connecting to the Oracle Database are obtained from a CSF map named `oracle.apps.security` with a key named `FUSION_APPS_ECSF_SES_ADMIN-KEY`.

Update the SEARCH_TOP and the Perl Command Locations used by Enterprise Scheduler Service (ESS) to Launch the SES Crawler

You must update the location for the Perl command, only if it has been changed after the recovery process. The default location for the Perl command is *ses_home*/perl/bin/perl.

To update the Perl command location:

1. Connect to the system where WebLogic Server middle tier is installed.

2. For Windows system, set the ESS application path by running the `essSetEnv.cmd` script present under the *ses_home*/bin directory.

3. Navigate to the *ses_home*/common/bin directory.

4. Start the WebLogic Server Administration Scripting Tool by executing the command `wlst.sh`.

5. Run the `connect()` command at the `wls/offline>` prompt. You will be prompted to enter the following values:

   - WebLogic Server user name

- WebLogic Server password

- WebLogic Server URL

6. Run the following command at the `wls:domain_name/serverConfig>` prompt to set the `SEARCH_TOP` location:

   ```
   wls:/>essManageRuntimeConfig("SearchEss","APP", operation="add",name="SEARCH_
   TOP",val="search_top_location")
   ```
   where, *search_top_location* is the absolute file path of the Oracle SES 11.2.2.2 home directory (that is, *ses_home* directory). For Windows system, use double backslashes (\\) as path delimiters.

7. Run the following command at the `wls:domain_name/serverConfig>` prompt to set the Perl command location:

   ```
   wls:/>essManageRuntimeConfig("SearchEss","ESS",
   operation="add",name="PerlCommand",val="perl_location")
   ```
   where, *perl_location* is the location for the Perl command in the Oracle SES 11.2.2.2 instance. For Windows system, use double backslashes (\\) as path delimiters.

Cold Backups

As an additional precaution to minimize downtime, you can perform a cold backup to backup all the data of an Oracle SES Instance. To back up an Oracle SES instance data, you must save a copy of the directories *oracle_base*, `oraInventory`, and `oradata`.

To perform a cold backup:

1. Shut down the Oracle SES instance (middle tier as well as database server).

2. Log in to the computer as the root user or the administrator.

3. Copy all the files under the Oracle Database directory *oracle_base*, the Oracle Inventory directory `oraInventory`, and the Oracle data storage directory `oradata`.

   These locations are specified during the Oracle SES installation. There are several ways to make a copy. For example, using the `tar` command:

   ```
   tar cvf ses_orabase.tar   {full path to Oracle base}
   tar cvf ses_orahome.tar   {full path to Oracle home}
   tar cvf ses_orainv.tar    {full path to oraInventory}
   tar cvf ses_oradat.tar    {full path to oradata}
   ```

   For example, if the `oradata` location is *oracle_base*/oradata, then save a copy using the command:

   ```
   tar cvf ses_oradat.tar oracle_base/oradata
   ```

4. Copy the `.tar` files created in steps 2, 3, and 4 to a safe location.

   > **Note:** You can use any compression method to perform file backup. For example, you can zip the files.

To recover files from a cold backup:

1. Shut down the Oracle SES instance (middle tier as well as Database server).

2. Restore all backed-up files. First decompress (`untar`) the files, then move them back to their original locations.

3. Start the Oracle SES instance.

Tuning Crawl Performance

Your Web crawling strategy can be as simple as identifying a few well-known sites that are likely to contain links to most of the other intranet sites in your organization. You could test this by crawling these sites without indexing them. After the initial crawl, you have a good idea of the hosts that exist in your intranet. You could then define separate Web sources to facilitate crawling and indexing on individual sites.

However, the process of discovering and crawling your organization's intranet, or the Internet, is generally an interactive one characterized by periodic analysis of crawling results and modification to crawling parameters. For example, if you observe that the crawler is spending days crawling one Web host, then you might want to exclude crawling at that host or limit the crawling depth.

This section contains the most common things to consider to improve crawl performance:

- Understanding the Crawler Schedule
- Registering a Proxy
- Checking Boundary Rules
- Checking Dynamic Pages
- Checking Crawler Depth
- Checking Robots Rule
- Checking Duplicate Documents
- Checking Redirected Pages
- Checking URL Looping
- Increasing the Oracle Redo Log File Size
- Increasing the Crawler Heap Size
- Increasing the Crawler Cache Queue Size
- Increasing the Number of Crawler Cache Threads
- What to Do Next

> **See Also:** "Monitoring the Crawling Process" on page 4-34 for more information on crawling parameters

Understanding the Crawler Schedule

Schedules define the frequency at which the Oracle SES index is updated with information about each source. This section describes characteristics the Oracle SES crawler schedule.

- The Failed Schedules section on the Home - General page lists all schedules that have failed. A failed schedule is one in which the crawler encountered an irrecoverable error, such as an indexing error or a source-specific login error, and cannot proceed. A failed schedule could be because of a partial collection and indexing of documents.
- The smallest granularity of the schedule interval is one hour. For example, you cannot start a schedule at 1:30 am.

- If a crawl takes longer to finish than the scheduled interval, then it starts again when the current crawl is done. Currently, there is no option to have the scheduled time automatically pushed back to the next scheduled time.

- When multiple sources are assigned to one schedule, the sources are crawled one by one following the order of their assignment in the schedule.

- The schedule starts crawling the assigned sources in the assigned order. Only one source is crawling under a schedule at any given time. If a source crawl fails, then the rest of the sources assigned after it are not crawled. The schedule does not restart. You must either resolve the cause of the failure and resume the schedule, or remove the failed source from the schedule.

- There is no automatic e-mail notification of schedule success or failure.

For more information about documents that the crawler does not index:

- Browse the crawler log in the Oracle SES Administration GUI. Select the Schedules subtab from the Home page, then click the Log File icon for the schedule.

- In the Oracle SES Administration GUI, select the **Statistics** subtab from the Home page. Under Crawler Statistics, choose **Problematic URLs**. This page lists errors encountered during the crawling process and the number of URLs that caused each error.

Registering a Proxy

By default, Oracle SES is configured to crawl Web sites in the intranet, so no additional configuration is required. However, to crawl Web sites on the Internet (also referred to as external Web sites), Oracle SES needs the HTTP proxy server information.

To register a proxy:

1. On the Global Settings page under Sources, select **Proxy Settings**.

2. Enter the proxy server name and port. Click **Set Proxy**.

3. Enter the internal host name suffix under Exceptions, so that internal Web sites do not go through the proxy server. Click **Set Domain Exceptions**.

   To exclude the entire domain, omit `http`, begin with `*.`, and use the suffix of the host name. For example, `*.us.example.com` or `*.example.com`. Entries without the `*.` prefix are treated as a single host. Use the IP address only when the URL crawled is also specified using the IP for the host name. They must be consistent.

4. If the proxy requires authentication, then enter the proxy authentication information on the Global Settings - Authentication page.

Checking Boundary Rules

The seed URL you enter when you create a source is turned into an inclusion rule. For example, if `www.example.com` is the seed URL, then Oracle SES creates an inclusion rule that only URLs containing the string `www.example.com` are crawled.

However, suppose that the example Web site includes URLs starting with `www.exa-mple.com` or `example.com` (without the `www`). Many pages have a prefix on the site name. For example, the investor section of the site has URLs that start with `investor.example.com`.

Always check the inclusion rules before crawling, then check the log after crawling to see what patterns have been excluded.

In this case, you might add `www.example.com`, `www.exa-mple.com`, and `investor.example.com` to the inclusion rules. Or you might just add `example`.

To crawl outside the seed site (for example, if you are crawling `text.us.xyz.com`, but you want to follow links outside of `text.us.xyz.com` to `xyz.com`), consider removing the inclusion rules completely. Do so carefully. This action could lead the crawler into many, many sites.

Notes for File Sources

- If no boundary rule is specified, then crawling is limited to the underlying file system access privileges. Files accessible from the specified seed file URL are crawled, subject to the default crawling depth. The depth, which is 2 by default, is set on the **Global Settings - Crawler Configuration** page. For example, if the seed is `file://localhost/home/user_a/`, then the crawl picks up all files and directories under `user_a` with access privileges. It crawls any documents in the directory `/home/user_a/level1` due to the depth limit. The documents in the `/home/user_a/level1/level2` directory are at `level 3`.

- The file URL can be in UNC (universal naming convention) format. The UNC file URL has the following format for files located within the host computer:

  `file://localhost///LocalComputerName/SharedFolderName`

  For example, specify `\\stcisfcr\docs\spec.htm` as `file://localhost///stcisfcr/docs/spec.htm`

  where `stcisfcr` is the name of the host computer.

  The string `localhost` is optional. You can specify the URL path without the string `localhost` in the URL, in which case the URL format is:

  `file:///LocalComputerName/SharedFolderName`

  For example,

  `file:///stcisfcr/docs/spec.htm`

  Note that you cannot use the UNC format to access files on other computers.

- On some computers, the path or file name could contain non-ASCII and multibyte characters. URLs are always represented using the ASCII character set. Non-ASCII characters are represented using the hex representation of their UTF-8 encoding. For example, a space is encoded as `%20`, and a multibyte character can be encoded as `%E3%81%82`.

  You can enter spaces in simple (not regular expression) boundary rules. Oracle SES automatically encodes these URL boundary rules. For example, `Home Alone` is specified internally as `Home%20Alone`. Oracle SES does this encoding for the following:

  - File source simple boundary rules

  - URL string tests

  - File source seed URLs

  Oracle SES does not alter regular expression rules. You must ensure that the regular expression rule specified is against the encoded file URL. Spaces are not allowed in regular expression rules.

Checking Dynamic Pages

Indexing dynamic pages can generate too many URLs. From the target Web site, manually navigate through a few pages to understand what boundary rules should be set to avoid crawling of duplicate pages.

Checking Crawler Depth

Setting the crawler depth very high (or unlimited) could lead the crawler into many sites. Without boundary rules, a crawler depth of 20 probably crawls the entire World Wide Web from most locations.

Checking Robots Rule

You can control which parts of your sites can be visited by robots. If robots exclusion is enabled (the default), then the Web crawler traverses the pages based on the access policy specified in the Web server `robots.txt` file.

The following sample `robots.txt` file specifies that no robots visit any URL starting with `/cyberworld/map/` or `/tmp/` or `/foo.html`:

```
# robots.txt for http://www.example.com/

User-agent: *
Disallow: /cyberworld/map/
Disallow: /tmp/
Disallow: /foo.html
```

If the Web site is under your control, then you can tailor a specific robots rule for the crawler by specifying Oracle Secure Enterprise Search as the user agent. For example:

```
User-agent: Oracle Secure Enterprise Search

Disallow: /tmp/
```

The robots `meta` tag can instruct the crawler either to index a Web page or to follow the links within it. For example:

```
<meta name="robots" content="noindex,nofollow">
```

Checking Duplicate Documents

Oracle SES always removes duplicate (identical) documents. Oracle SES does not index a page that is identical to one it has already indexed. Oracle SES also does not index a page that it reached through a URL that it has already processed.

With the Web Services API, you can enable or disable *near* duplicate detection and removal from the result list. Near duplicate documents are similar to each other. They may or may not be identical to each other.

> **See Also:** "Oracle Secure Enterprise Search Web Services APIs" on page 13-2

Checking Redirected Pages

The crawler crawls only redirected pages. For example, a Web site might have Javascript that redirects users to another site with the same title. In such cases, only the redirected site is indexed.

Check for inclusion rules from redirects. The inclusion rules are based on the type of redirect. The `EQ_TEST.EQ$URL` table stores all of the URLs that have been crawled or are scheduled to be crawled. There are three kinds of redirects defined in it:

- **Temporary Redirect**: A redirected URL is always allowed if it is a temporary redirection (HTTP status code 302, 307). Temporary redirection is used for whatever reason that the original URL should still be used in the future. It's not possible to find out temporary redirect from EQ$URL table other than filtering out the rest from the log file.

- **Permanent Redirect**: For permanent redirection (HTTP status 301), the redirected URL is subject to boundary rules. Permanent redirection means the original URL is no longer valid and the user should start using the new (redirected) one. In EQ$URL, HTTP permanent redirect has the status code 954

- **Meta Redirect**: Metatag redirection is treated as a permanent redirect. Meta redirect has status code 954. This is always checked against boundary rules.

The `STATUS` column of `EQ_TEST.EQ$URL` lists the status codes. For descriptions of the codes, refer to Appendix B, "URL Crawler Status Codes."

> **Note:** Some browsers, such as Mozilla and Firefox, do not allow redirecting a page to load a network file. Microsoft Internet Explorer does not have this limitation.

Checking URL Looping

URL looping refers to the scenario where a large number of unique URLs all point to the same document. Looping sometimes occurs where a site contains a large number of pages, and each page contains links to every other page in the site. Ordinarily this is not a problem, because the crawler eventually analyzes all documents in the site. However, some Web servers attach parameters to generated URLs to track information across requests. Such Web servers might generate a large number of unique URLs that all point to the same document.

For example,

`http://example.com/somedocument.html?p_origin_page=10`

might refer to the same document as

`http://example.com/somedocument.html?p_origin_page=13`

but the `p_origin_page` parameter is different for each link, because the referring pages are different. If a large number of parameters are specified and if the number of referring links is large, then a single unique document could have thousands or tens of thousands of links referring to it. This is an example of how URL looping can occur.

Monitor the crawler statistics in the Oracle SES Administration GUI to determine which URLs and Web servers are being crawled the most. If you observe an inordinately large number of URL accesses to a particular site or URL, then you might do one of the following:

- Exclude the Web server: This prevents the crawler from crawling any URLs at that host. (You cannot limit the exclusion to a specific port on a host.)

- Reduce the crawling depth: This limits the number of levels of referred links the crawler follows. If you are observing URL looping effects on a particular host, then you should take a visual survey of the site to find out an estimate of the depth of the leaf pages at that site. Leaf pages are pages that do not have any links to other

pages. As a general guideline, add three to the leaf page depth, and set the crawling depth to this value.

Be sure to restart the crawler after altering any parameters. Your changes take effect only after restarting the crawler.

Increasing the Oracle Redo Log File Size

Oracle SES allocates 200MB of disk space for the redo log during installation, which is sufficient to crawl a relatively large number of documents, such as around 300GB of text. If you are going to install Oracle SES application on an existing Oracle Database, then make sure that at least 200MB of disk space is allocated for the redo log.

However, if your disk has sufficient space to increase the redo log, and if you are going to crawl a very large number of documents, for example, more than 300GB of text, then increase the redo log file size for better crawl performance.

> **Note:** The biggest transaction during crawling is SYNC INDEX by Oracle Text. Check the AWR report or the V$SYSSTAT view to see the actual redo size during crawling. Roughly, 200MB is sufficient to crawl up to 300GB.

To increase the size of the redo log files:

1. Open SQL*Plus and connect as the SYSTEM user. It has the same password as that of the SEARCHSYS user.

2. Issue the following SQL statement to see the current redo log status:

```
SELECT vl.group#, member, bytes, vl.status
    FROM v$log vl, v$logfile vlf
    WHERE vl.group#=vlf.group#;

GROUP# MEMBER                                              BYTES STATUS
------ -------------------------------------------------- ---------- ----------
     3 /scratch/ses111/oradata/o11101/redo03.log       209715200 INACTIVE
     2 /scratch/ses111/oradata/o11101/redo02.log       209715200 CURRENT
     1 /scratch/ses111/oradata/o11101/redo01.log       209715200 INACTIVE
```

3. Drop the INACTIVE redo log file. For example, to drop group 3:

```
ALTER DATABASE DROP LOGFILE group 3;

Database altered.
```

4. Create a larger redo log file with a command like the following. If you want to change the file location, specify the new location.

```
ALTER DATABASE ADD LOGFILE '/scratch/ses111/oradata/o11101/redo03.log' 2
    size 400M reuse;
```

5. Check the status to ensure that the file was created.

```
SELECT vl.group#, member, bytes, vl.status
    FROM v$log vl, v$logfile vlf
    WHERE vl.group#=vlf.group#;

GROUP# MEMBER                                              BYTES STATUS
------ -------------------------------------------------- ---------- ----------
     3 /scratch/ses111/oradata/o11101/redo03.log       419430400 UNUSED
```

```
    2 /scratch/ses111/oradata/o11101/redo02.log              209715200 CURRENT
    1 /scratch/ses111/oradata/o11101/redo01.log              209715200 INACTIVE
```

6. To drop a log file with a CURRENT status, issue the following ALTER statement, then check the results.

```
ALTER SYSTEM SWITCH LOGFILE;

SELECT vl.group#, member, bytes, vl.status
    FROM v$log vl, v$logfile vlf
    WHERE vl.group#=vlf.group#;

GROUP# MEMBER                                                 BYTES STATUS
------ ------------------------------------------------- ---------- ----------
    3 /scratch/ses111/oradata/o11101/redo03.log          419430400 CURRENT
    2 /scratch/ses111/oradata/o11101/redo02.log          209715200 ACTIVE
    1 /scratch/ses111/oradata/o11101/redo01.log          209715200 INACTIVE
```

7. Issue the following SQL statement to change the status of Group 2 from ACTIVE to INACTIVE:

```
ALTER SYSTEM CHECKPOINT;

SELECT vl.group#, member, bytes,  vl.status
    FROM v$log vl, v$logfile vlf
    WHERE vl.group#=vlf.group#;

GROUP# MEMBER                                                 BYTES STATUS
------ ------------------------------------------------- ---------- ----------
    3 /scratch/ses111/oradata/o11101/redo03.log          419430400 CURRENT
    2 /scratch/ses111/oradata/o11101/redo02.log          209715200 INACTIVE
    1 /scratch/ses111/oradata/o11101/redo01.log          209715200 INACTIVE
```

8. Repeat steps 3, 4 and 5 for redo log groups 1 and 2.

Increasing the Crawler Heap Size

You can specify a particular value for crawler heap size by updating the value of -mx parameter of the CRAWLER_EXEC_PARAMS configuration setting in the *ses_home*/bin/clexecutor.sh file on a Linux or UNIX system, and the *ses_home*\bin\clexecutor.cmd file on a Windows system. For example, to specify heap size of 1GB, set the -mx parameter value to -mx1024m.

> **Note:** There is no need to restart Oracle SES server after changing the heap size in the clexecutor script file and the subsequent crawl processes will start using the updated heap size.

Increasing the Crawler Cache Queue Size

If the system on which Oracle SES instance middle-tier is installed has sufficiently large memory, then you can increase the size of the crawler cache queue, which is an in-memory queue used by the crawler to process the documents. The following are the steps to increase the size of crawler cache queue:

1. Export the current crawler configuration settings using the searchadmin command. For example:

```
searchadmin -c http://ses_server_host:ses_server_
port/search/api/admin/AdminService -p ses_admin_password export crawlerSettings
```

```
-o crawlerSettings.xml
```

2. Open the `crawlerSettings.xml` file and set the required values for `minCacheQueue` and `maxCacheQueue` elements. The `minCacheQueue` element represents the minimum size for the crawler cache queue and `maxCacheQueue` element represents the maximum size for the crawler cache queue. These element values are in megabytes. For example, the following configuration specifies the minimum crawler cache queue size of 2MB and maximum crawler cache queue size of 100MB.

```
<search:minCacheQueue>2</search:minCacheQueue>
<search:maxCacheQueue>100</search:maxCacheQueue>
```

> **Note:** The maximum crawler cache queue size limit is 100MB.

3. Update the changes using the `searchadmin` command. For example:

```
searchadmin -c http://ses_server_host:ses_server_
port/search/api/admin/AdminService -p ses_admin_password update crawlerSettings
--UPDATE_METHOD=overwrite -i crawlerSettings.xml
```

> **Note:** The minimum crawler cache queue size and maximum crawler cache queue size configuration settings are also logged in the crawler log file as shown in the following example:
>
> ```
> 2014-02-19 14:26:50.754 NOTIFICATION   Main Thread   EQG-30527
> Caching queue high water mark = 100 MB
> 2014-02-19 14:26:50.754 NOTIFICATION   Main Thread   EQG-30528
> Caching queue low water mark = 2 MB
> ```

Increasing the Number of Crawler Cache Threads

If the system on which Oracle SES instance middle-tier is installed has a large number of CPUs, then the number of caching threads that are used to save documents to the database can be increased by updating the value of `cachingThreads` element of the crawler configuration XML file that is exported using the `searchadmin` command.

For example, the following configuration setting specifies the number of caching threads to be 5.

```
<search:cachingThreads auto="false">
  <search:numThreads>5</search:numThreads>
</search:cachingThreads>
```

> **Note:** Setting the number of caching threads to a very high value may cause contention of Oracle SES database resources. Oracle recommends that you do not increase the number of caching threads beyond 8.

What to Do Next

If you are still not crawling all the pages you think you should, then check which pages were crawled by doing one of the following:

- Check the crawler log file
- Create a search source group

To check the crawler log file:

1. On the home page, click the **Schedules** secondary tab to display the Crawler Schedules page.

2. Click the Log File icon to display the log file for the source.

3. To obtain the location of the full log, click the **Status** link. The Crawler Progress Summary and Log Files by Source section displays the full path to the log file.

To create a search source group:

1. On the Search page, click the **Source Groups** subtab.

2. Click **New** to display Create New Source Group Step 1.

3. Enter a name, then click **Proceed to Step 2**.

4. Select a source type, then shuttle only one source from Available Sources to Assigned Sources.

5. Click **Finish**.

To search the source group:

1. On any page, click the **Search** link in the top right corner to open the query application.

2. Select the group name, then issue a search term to list the matches within the source.

3. Select the group name, then click **Browse** to see a list of search groups:

   - The number after the group name identifies the number of browsed documents. Click the number to browse the search results.

   - Click the arrow before the group name to display a hierarchy of search results. The number of matches appears after each item in the hierarchy.

Tuning Search Performance

Oracle SES contains a lot of features that optimize the search performance. This section contains suggestions on how to improve the response time and throughput performance of Oracle SES. It identifies the most common ways to improve search quality.

- Adding Suggested Links

- Authentication and Authorization

- Parallel Querying and Index Partitioning

- Managing Index Fragmentation

- Optimizing the Index

- Adjusting the Indexing Parameters

- Checking the Search Statistics

- Load Balancing on Oracle RAC

- WebLogic Search Server Configuration

- Database Initialization Parameters

- Relevancy Boosting

- [Increasing the JVM Heap Size](#)

- [Increasing the Oracle Undo Space](#)

> **See Also:** ["Searching on Date Attributes"](#) on page 3-9

Adding Suggested Links

Suggested links enable you to direct users to a designated Web site for particular query keywords. For example, when users search for "Oracle Secure Enterprise Search documentation" or "Enterprise Search documentation" or "Search documentation", you could suggest `http://www.oracle.com/technology`.

Suggested link keywords are rules that determine which suggested links are returned (as suggestions) for a query. A rule can include query terms and logical operators. For example, "secure AND search". With this rule, the corresponding suggested link is returned for the query "secure enterprise search", but it is not returned for the query "secure database".

The rule language used for the indexed queries supports the following operators:

Table 12–1   Suggested Link Keyword Operators

Operator	Example
ABOUT	about(dogs)
AND	dog and cat
NEAR	dog ; cat
OR	dog or cat
PHRASE	dog sled
STEM	$dog
THESAURUS	SYN(dog)

> **Note:** Do not use special characters, such as #, $, =, and &, in keywords.

Suggested links appear at the top of the search result list. Oracle SES can display up to two suggested links for each query.

This feature is especially useful for providing links to important Web pages that are not crawled by Oracle Secure Enterprise Search. Add or edit suggested links on the Search - Suggested Links page in the Oracle SES Administration GUI.

Authentication and Authorization

By tuning the security filter settings, you can prevent time outs.

To change the configuration of the security filter:

1. Log in to the Oracle SES Administration GUI.

2. Click the Global Settings tab, then Query Configuration.

3. Scroll down to Security Filter Configuration and change these settings using the guidelines provided in the Help.

   - Security Filter Lifespan

- Authentication Timeout

- Authorization Timeout

- Minimum Number of Threads

- Maximum Number of Threads

You can further tune the security filter by using the Administration API to set the `<search:preserveStaleSecurityFilterOnError>` and `<search:securityFilterRefreshWaitTimeout>` parameters in the `queryConfig` object. For example, these settings allow an expired security filter to be used immediately when a fresh security filter is unavailable:

```
<search:preserveStaleSecurityFilterOnError>true
    </search:preserveStaleSecurityFilterOnError>
<search:securityFilterRefreshWaitTimeout>0
    </search:securityFilterRefreshWaitTimeout>
```

The settings listed previously are also parameters of the `queryConfig` object and can be modified using the API. See the *Oracle Secure Enterprise Search Administration API Guide*.

Parallel Querying and Index Partitioning

Parallel querying significantly improves search performance and facilitates searches of very large data sources. You can optimize query performance of large document sources by storing the crawler index in partitions distributed across several independent disks. Oracle SES then executes parallel sub-queries automatically against the partitions. Both I/O and CPU resources are used in parallel. The parallel query architecture is based on Oracle Database partitioning and enhancements in Oracle Text.

The Parallel Query feature is automatically supported by Oracle SES when the partitioning option is enabled. You can specify partitioning only during the Oracle SES installation. Partitions can be configured in Oracle SES only using the Administration API.

The default tablespaces for Oracle SES are SEARCH_DATA, SEARCH_INDEX, and SEARCH_TEMP, which can be used for partitioning.

> **Note:** To make the best use of the Parallel Query feature, Oracle recommends that you run Oracle SES on a server with a 4-Core CPU, with at least 8 GB of RAM and multiple fast disk drives.

To enable partitioning:

1. Acquire a license for the Oracle Partitioning option for Oracle Database.

2. During the Oracle SES software installation, select the Oracle Partitioning Option setting. Oracle Database is then installed with the partitioning feature.

3. Activate and configure partitions using Administration API.

To use partitioned tablespaces in Oracle SES:

1. Create one or more ASSM (Automatic Segment Space Management) tablespaces using a tool such as Enterprise Manager.

2. Open a `searchadmin` interactive session.

> **See Also:** "Opening an Interactive Session" in *Oracle Secure Enterprise Search Administration API Guide*

3. Activate the `partitionConfig` object.

> **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for more information about the `partitionConfig` object type.

4. Register storage areas - Update the `storageArea` object to register the existing tablespaces as storage areas for use by Oracle SES. To make optimum use of the parallel query feature, you should distribute partitioned tablespaces across all the physical disks.

> **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for more information about the `storageArea` object type.

5. Configure partitions - Update the `partitionConfig` object with the above `storageArea` objects, partition attributes, and partition rules. A partition typically includes multiple storage areas.

6. Create data sources and schedule them for crawling.

Example: Adding a Tablespace and Using it in a Partition Rule

This example registers a new tablespace for use by Oracle SES:

1. Create a new ASSM tablespace. This example uses SQL to create a tablespace named NEW_ONE:

```
CREATE TABLESPACE new_one DATAFILE '/ses_storage/new_one.dbf'
   SIZE 8G REUSE AUTOEXTEND ON NEXT 2G MAXSIZE UNLIMITED
   EXTENT MANAGEMENT LOCAL SEGMENT SPACE MANAGEMENT AUTO;
```

2. Open a `searchadmin` interactive session:

```
$ searchadmin --CONNECTION=http://ses_host:ses_
port/search/api/admin/AdminService
```

3. Activate the `partitionConfig` object:

```
activate partitionConfig
```

4. Export the XML description of the partition configuration to a file named `part.xml`:

```
export partitionConfig --OUTPUT_FILE=part.xml
```

5. Create an XML file named `search_data.xml` and describe the NEW_ONE tablespace as an Oracle SES storage area, as shown here:

```
<?xml version="1.0" encoding="UTF-8"?>
<search:config xmlns:search="http://xmlns.oracle.com/search"
productVersion="11.2.2.2.0">
   <search:storageAreas>
      <search:storageArea>
         <search:name>NEW_ONE</search:name>
         <search:description>Additional storage area</search:description>
         <search:usage>PARTITION</search:usage>
      </search:storageArea>
   </search:storageAreas>
</search:config>
```

6. Open `part.xml` in a text editor and edit the `<search:ruleType>` and `<search:storageArea>` elements as shown here. This example hashes all documents into two partitions: one partition in the SEARCH_DATA tablespace, and the other partition in the NEW_ONE tablespace.

```
<?xml version="1.0" encoding="UTF-8"?>
<search:config xmlns:search="http://xmlns.oracle.com/search"
productVersion="11.2.2.2.0">
    <search:partitionConfig>
        <search:partitionRules>
            <search:partitionRule>
                <search:partitionValue>EQ_DEFAULT</search:partitionValue>
                <search:valueType>META</search:valueType>
                <search:ruleType>HASH</search:ruleType>
                <search:storageArea>SEARCH_DATA,NEW_ONE</search:storageArea>
            </search:partitionRule>
        </search:partitionRules>
    </search:partitionConfig>
</search:config>
```

7. Register the new storage area:

```
create storageArea --NAME=new_one --INPUT_FILE=search_data.xml
```

8. Update the partition configuration:

```
update partitionConfig --INPUT_FILE=part.xml --UPDATE_METHOD=overwrite
```

Limitations of Configuring Partitions

The following are some of the limitations while configuring a partition:

- Once the partition option is enabled, then it cannot be disabled, and hence, parallel querying also cannot be disabled.

- A partition cannot be dropped if it is in use.

- When a data source is moved from one partition to another, the data source must be recrawled for reindexing purpose.

- You can use one of the following methods for updating a partition rule:

  - Update the rule and then recrawl the data source.

  - Delete the data source, delete and recreate the rule, recreate the data source, and recrawl the data source.

Managing Index Fragmentation

Index fragmentation management allows the search engine index to be updated while Oracle SES is executing searches. This is achieved by temporarily saving index changes to an in-memory index and periodically merging them with the larger disk-based search engine index. This reduces fragmentation and leads to faster response times. Index fragmentation management is implemented automatically on Oracle SES, but it can be tuned by configuring Oracle Text, where you can turn index fragmentation management on and off, and specify the frequency of index merges.

Optimizing the index also reduces fragmentation, and it can significantly increase the speed of searches. Schedule index optimization on a regular basis. Also, optimize the index after the crawler has made substantial updates or if fragmentation is more than

50%. Verify that index optimization is scheduled during off-peak hours. Optimization of a very large index could take several hours.

You can see the fragmentation level and run index optimization on the Global Settings - Index Optimization page in the Oracle SES Administration GUI. Index optimization has these options:

Do Not Run Optimization Longer Than
Specify a maximum duration for the index optimization process. The actual time taken for optimization does not exceed this limit, but it can be shorter. A longer optimization time results in a more optimized index. In this mode, the optimization process does not require a large amount of free disk space.

Until the Optimization is Finished
Specifies that the optimization continues until it is finished. Allowing the optimization to complete creates a more compact index and supports better performance than a partial optimization.

In this mode, Oracle SES creates a temporary copy of the index. The required disk space almost equals the current index size. If sufficient free disk space is not available, then the optimization fails. Use the appropriate SQL query shown here to estimate the minimum disk requirement:

- **Oracle SES Without Partitioning**

```
SELECT SUM(bytes)/1048576 AS "MBytes"
   FROM dba_segments
   WHERE segment_name IN ('DR$EQ$DOC_PATH_IDX$I','DR$EQ$DOC_PATH_IDX$X');
```

- **Oracle SES With Partitioning**

```
SELECT SUM(sz) AS "MBytes"
   FROM
   (
      SELECT MAX(bytes)/1048576 sz FROM dba_segments
         WHERE segment_name LIKE 'DR#EQ$DOC_PATH_IDX$%I'
   UNION
      SELECT MAX(bytes)/1048576 sz FROM dba_segments
         WHERE segment_name LIKE 'DR#EQ$DOC_PATH_IDX$%X'
   ) ;
```

These queries return an estimate of the *minimum* disk space needed for optimization. Oracle SES may require more disk space than this estimate.

After the optimization is complete, Oracle SES releases the disk space consumed during the optimization. The space can be used by future crawls or any activity that consumes disk space.

Optimizing the Index

Optimizing the **index** reduces fragmentation, and it can significantly increase the speed of searches. Schedule index optimization on a regular basis. Also, optimize the index after the crawler has made substantial updates or if fragmentation is more than 50%. Verify that index optimization is scheduled during off-peak hours. Optimization of a very large index could take several hours.

You can see the fragmentation level and run index optimization on the Global Settings - Index Optimization page in the Oracle SES Administration GUI. You can specify a maximum number of hours for the optimization to run, but for best performance, run

the optimization until completion. Oracle SES uses a faster optimization method and creates a more compact copy of the index when no time limit is set.

Adjusting the Indexing Parameters

To improve indexing performance, adjust the following parameters on the Global Settings - Set Indexing Parameters page of the Oracle SES Administration GUI:

Indexing Batch Size

When the crawled data in the cache directory reaches Indexing Batch Size, Oracle SES starts indexing. The bigger the batch size, the longer it takes to start indexing each batch. Only indexed data can be searched: Data in the cache cannot be searched. The default size is 250M.

Document fetching and indexing run concurrently. While indexing is running, the Oracle SES crawler continues to fetch documents and store them in the cache directory.

Indexing Memory Size

This is the upper limit of memory used for indexing before flushing the index to disk.

A large amount of memory improves indexing performance because it reduces I/O. It also improves query performance because the created index is less fragmented from the beginning, while a fragmented index can be optimized later. Set this parameter as high as possible without causing memory paging.

A smaller amount of memory might be useful when indexing progress should be tracked or when run-time memory is scarce. The default size is 275M. In general, increasing the Indexing Memory Size parameter can reduce fragmentation.

Parallel Indexing Degree

The number of concurrent threads used for indexing. This parameter is disabled in the current version of Oracle SES; it is always set to 1.

Checking the Search Statistics

See the Home - Statistics page in the Oracle SES Administration GUI for lists of the most popular queries, failed queries, and ineffective queries. This information can lead to the following actions:

- Refer users to a particular Web site for failed queries on the Search - Suggested Links page.

- Fix common errors that users make in searching on the Search - Alternate Words page.

- Make important documents easier to find on the Search - Relevancy Boosting page.

Every hour, Oracle SES automatically summarizes logged queries. The summarizing task might utilize the server resource if there are a large number of logged queries, and this might impact the query performance. This issue is visible for stress tests where several queries are executed every second. The ideal solution in such instances is to disable the query statistics option.

To disable the query statistics option:

1. On the home page, click **Global Settings**, then **Query Configuration**.

2. In the Query Statistics section, select **No** for the **Enable Query Statistics** option.

Cleaning up Search statistics data from the Database

Oracle SES stores the query log for seven days in the Database. If you want to clean up the query log from the Database immediately, then run the following SQL commands:

```
SQL> truncate table searchsys.eq$statistic;
SQL> truncate table searchsys.eq$user_clicks;
SQL> truncate table searchsys.EQ$SUM_CLICK_THROUGH;
SQL> truncate table searchsys.EQ$SUM_QUERY_STAT;
SQL> truncate table searchsys.EQ$SUM_STAT_DAILY;
SQL> truncate table searchsys.EQ$SUM_STAT_FAILED;
SQL> truncate table searchsys.EQ$SUM_STAT_INEFFECT;
SQL> truncate table searchsys.EQ$SUM_STAT_POPULAR;
```

Load Balancing on Oracle RAC

When Oracle SES is deployed in an Oracle Real Applications Cluster (Oracle RAC) environment, the usage profile is typically one of the following:

- Small index with a large query load

- Large index with a small-to-large query load

A third option, a small index and a small query load, typically operates on a single computer.

Configuring a Small Index

The load balancing solutions provided by Oracle RAC and the WebLogic Server are sufficient for this type of Oracle SES deployment. Most or all of the index can reside in memory or the buffer cache. You only need to set up the listeners appropriately for Oracle SES.

To set up the listeners:

- Provide a local listener on each Oracle RAC instance.

- Do not configure remote listeners.

- Oracle recommends dedicated processes over shared processes.

Oracle RAC Deployment Best Practice

When Oracle SES is deployed in an Oracle RAC environment, then do the configuration described in this section for the best query performance.

> **Note:** It is recommended to do this configuration if the AWR report shows any of the following statistics:
>
> - The Oracle RAC statistic for the estimated interconnect traffic shows a very high value.
>
> - The Oracle RAC cluster *wait* events are in the top 5 timed foreground events.
>
> - The Oracle RAC cluster *wait* class consumes a very large amount of database time.

Three Node Oracle RAC Environment

In a three node Oracle RAC environment, use one Oracle RAC instance for the crawler and use the rest of the Oracle RAC instances for the query application. This deployment configuration will increase the query performance due to separate Oracle

RAC nodes for query application and crawler. It will also reduce the Oracle RAC cache fusion overhead for the query application without compromising the high availability.

The above described configuration can be done by creating two database services - one for the crawler and one for the query application. After creating these database services, configure the database connection details for both these services.

> **See Also:**
>
> - "Update the WebLogic Server Connection Pool used by Oracle SES" on page 12-2 to configure the database connection details for the query application.
>
> - "Update the Database Connection String in the Credential Storage Framework (CSF) used by the Crawler" on page 12-2 to configure the database connection details for the crawler.

WebLogic Search Server Configuration

Oracle SES is installed in a WebLogic domain. The default settings for stuck threads can result in slow query performance even under a moderate load.

To change the Search server configuration

1. Log in to the WebLogic Administration Console.

2. In the left panel under Change Center, click **Lock & Edit**.

3. In the left panel under Domain Structure, expand Environment and click **Servers**. The Summary of Services page is displayed in the main panel.

4. In the Name column, click **search_server1**. The Settings for search_server1 page is displayed.

5. Select the **Configuration** tab.

6. Configure these settings:

   - Stuck Thread Max Time: 3600

   - Stuck Thread Timer Interval: 1800

7. Click **Save**.

8. Repeat these steps for any other search server instances, such as search_server2.

9. In the left panel under Change Center, click **Activate Changes**.

Database Initialization Parameters

To support a large number of simultaneous users, you may need to increase the values of these database initialization parameters:

- `PROCESSES`

- `SESSIONS`

- `OPEN_CURSORS`

The crawler also uses several threads, and each thread uses several database connections. You can alter the number of crawler threads on the Home - Sources - Crawling Parameters page of the Oracle SES Administration GUI.

Use the combined estimate of concurrent user processes and crawler threads for the value of `PROCESSES`. Then modify `SESSIONS` to a compatible value, typically calculated as `1.1 * PROCESSES`.

You can monitor the number of open cursors using the statistics stored in the V$SESSTAT dynamic performance view. If the number of open cursors for user sessions frequently approaches the maximum, then you can increase that number.

> **See Also:**
>
> - *Oracle Database Reference* for initialization parameters and dynamic performance views
>
> - *Oracle Database Performance Tuning Guide* for database tuning

To change the database initialization parameters:

1. Open SQL*Plus and log in to Oracle Database as a privileged user, such as SYSTEM.

2. For a list of all initialization parameters and their current settings, issue this SQL*Plus command:

   ```
   show parameters
   ```

3. Issue ALTER SYSTEM commands, using values appropriate for your system, to change the value of the parameters. For example, this command sets PROCESSES to 800:

   ```
   ALTER SYSTEM SET processes=800 SCOPE=spfile;
   ```

4. Restart Oracle Database for the new settings to take effect.

Relevancy Boosting

Relevancy boosting lets administrators influence the order of documents in the result list for a particular search. You might want to override the default results for the following reasons:

- For a highly popular search, direct users to the best results

- For a search that returns no results, direct users to some results

- For a search that has no click-throughs, direct users to better results

In a search, each result is assigned a score that indicates how relevant the result is to the search; that is, how good a result it is. Sometimes you know the documents that are highly relevant to some search. For example, your company Web site could have a home page for XML (http://example.com/XML-is-great.htm), which you want to appear high in the results of any search for XML. You would boost the score of the XML home page to 100 for an XML search.

The document also has a score computed for searches that are not among the boosted queries.

Two methods can help you locate URLs for relevancy boosting: *locate by search* and *manual URL entry*.

Relevancy boosting, like end user searching, is case-insensitve. For example, a document with a boosted score for Oracle is boosted for oracle.

Increasing the JVM Heap Size

If you expect heavy loads on the Oracle SES server, then configure the Java Virtual Machine (JVM) heap size for better performance.

You can configure the JVM heap size using the WebLogic Server JVM configuration parameters. Refer to *Oracle Fusion Middleware Performance and Tuning for Oracle WebLogic Server* for more information.

> **Note:** The following are some of the important points to consider while configuring the JVM heap size:
>
> - The heap size should be set to about 50% of the generally consumed heap size under normal application load.
>
> - The heap size should not be too large, for example, its value should not be more than 2 GB. Setting the heap size to a very large value may severely impact the application performance due to its usage by the Java Garbage Collection (GC) process.

Increasing the Oracle Undo Space

Heavy query load should not coincide with heavy crawl activity, especially when there are large-scale changes on the target site. If it does, such as when a crawl is scheduled around the clock, then increase the size of the Oracle undo tablespace with the `UNDO_RETENTION` parameter.

> **See Also:** *Oracle Database SQL Language Reference* and *Oracle Database Administrator's Guide* on Oracle Technology Network for more information about increasing the Oracle undo space

Optimizing Query Application Performance

If you plan to use the Oracle SES default query user interface and have an Oracle Application Server Web Cache installation, then you can use its compression utility to compress the content Oracle SES sends over the network. For example, the utility can compress `results.jsp` from 980 to 72K. Compression provides the greatest benefit to users connecting over the Internet.

Use these Web cache compression rules:

```
/search/search?(.*)
/search/results.jsp?(.*)
```
OracleAS Web Cache does not benefit custom querying applications.

Using Command Line Tools

You can use the `searchctl` command line tool for starting and stopping the Oracle SES instance, only when both the Database and the WebLogic Server middle tier are installed as part of the Oracle SES software installation.

For more information about various `searchctl` command parameters, refer to the section "Using `searchctl` Command" on page 7.

Configuring Oracle Diagnostic Logging (ODL) for Oracle SES Server

You can configure Oracle Diagnostic Logging (ODL) for Oracle SES instance using the Oracle Enterprise Manager (EM) as well as using the WebLogic Server Administration Scripting Tool **wlst.sh**.

> **Note:** The ODL logging configuration described here is for Oracle SES managed server and is not for Oracle SES crawler. For information about Oracle SES crawler log, refer to the section "Viewing Crawler Logs" on page 4-35.

To configure ODL using Oracle Enterprise Manager:

1. Log on to the Oracle Enterprise Manager UI using the following URL:

   `http://host:port/em`
   where, *host* and *port* are the host name and the port for the WebLogic Server middle tier.

2. On the Login page, provide user name and password as `weblogic` and the Oracle SES administrator password respectively.

3. In the left panel, from the **Farm_*base_domain*** tree view, navigate to **WebLogic Domain** > *domain_name* > *ses_managed_server*. The value of *domain_name* is `base_domain` by default, and the value of *ses_managed_server* is `search_server1` by default.

4. Click the **WebLogic Server** list box, and navigate to **Logs > Log Configuration**.

5. Under Logger Name, navigate to the tree node **Root Logger > oracle > oracle.search**.

6. Select the required logging level from the Oracle Diagnostic Logging Level list.

7. Click **Apply** to save the changes.

To configure ODL using the WebLogic Server Administration Scripting Tool wlst.sh:

1. Connect to the Linux system where WebLogic Server middle tier is installed.

2. Go to `MW_HOME/oracle_common/common/bin`.

3. Run the `wlst.sh` command.

4. At the `wls/offline>` prompt, run the following command:

   `connect('weblogic_username','weblogic_password','t3://localhost:7001')`

5. Run the following command to set the required log level:

   `setLogLevel(target="ses_managed_server",logger="oracle.search", level="ODL_log_level")`
   The *ODL_log_level* can have one of the following values:

   - `SEVERE`

   - `WARNING`

   - `INFO`

   - `CONFIG`

   - `FINE`

   - `FINER`

   - `FINEST`

Viewing Oracle SES Server Log Files

You can troubleshoot the run-time issues related to Oracle SES by viewing the server logs associated with the Oracle SES instance.

Oracle SES server log files

Refer to the Oracle SES server log files in case you encounter problems such as:

- Oracle SES application hangs, that is, it does not respond to user actions
- Oracle SES Administration GUI pages do not load
- Oracle SES application throws database exception

The Oracle SES server log files are stored in the directory: *wls_domain_home*/*ses_domain_name*/servers/search_server1/logs.

Oracle Enterprise Scheduler server log files

Refer to the Oracle Enterprise Scheduler server log files in case you encounter problems such as:

- Crawler fails to launch
- Oracle SES Administration GUI pages related to schedules do not load

The Oracle Enterprise Scheduler server log files are stored in the directory: *wls_domain_home*/*ses_domain_name*/servers/ess_server1/logs.

> **Note:** You can view all the stuck threads on the WebLogic Console at **Servers >** *server name* **> Monitoring > Threads**, where *server name* is **search_server1** for Oracle SES server and **ess_server1** for Oracle Enterprise Scheduler server.

> **See Also:**
>
> - "Configuring Oracle Diagnostic Logging (ODL) for Oracle SES Server" on page 12-23
> - "Viewing Crawler Logs" on page 4-35
> - "Monitoring Oracle SES Instance" on page 12-25

Monitoring Oracle SES Instance

In a production environment, where a load balancer or other monitoring tools are used to ensure system availability, Oracle SES instance can be monitored easily using the following URL:

```
http://host:port/monitor/check.jsp
```

The page should display the following message: `Oracle Secure Enterprise Search instance is up.`

This message is not translated to other languages because system monitoring tools might need to byte-compare this string.

If Oracle SES is not available, then the page displays either a connection error or the HTTP status code 503.

Integrating Oracle SES with Google Desktop

Oracle Secure Enterprise Search provides a GDfE plug-in to integrate with Google Desktop Enterprise Edition. You can include Google Desktop results in your Oracle SES hit list. You can also link to Oracle SES from the GDfE interface.

> **See Also:** Google Desktop for Enterprise Plug-in Readme at
> http://*host:port*/search/query/gdfe/gdfe_readme.html

Accessing Oracle WebLogic Server Administration Console

The Oracle WebLogic Server Administration Console is a Web browser-based user interface that displays the current status of the Oracle SES middle tier.

The Application Server Control Console is installed and configured automatically with WebLogic Server and can be accessed when the WebLogic Server is running.

> **See Also:** Starting and Stopping Oracle SES Instance

To access the Oracle WebLogic Server Administration Console:

1. Enter the following URL in a Web browser, replacing *host:port* with the host name and the port for the WebLogic Administration Console:

   http://*host:port*/console

2. Log in as the weblogic user with your Oracle SES administrator password.

   > **See Also:** http://docs.oracle.com/cd/E23943_
   > 01/apirefs.1111/e13952/core/index.html
   >
   > for detailed documentation related to the Oracle WebLogic Server
   > Administration Console.

13

Oracle Secure Enterprise Search APIs

This chapter explains the Oracle Secure Enterprise Search (Oracle SES) APIs and related information. This chapter contains the following topics:

- Overview of Oracle Secure Enterprise Search APIs
- Oracle Secure Enterprise Search Web Services APIs
- Oracle Secure Enterprise Search Java SDK

> **See Also:** *Oracle Secure Enterprise Search Java API Reference*

Overview of Oracle Secure Enterprise Search APIs

Oracle Secure Enterprise Search provides the following APIs:

Web Services APIs

The Web Services APIs are used to integrate Oracle SES search capabilities into your search application. Oracle SES provides Java proxy libraries. You either can use the Java libraries or create proxies, based on the published Web Services Description Language (WSDL) files, to access Oracle SES Web Services. Oracle SES provides two Web Services APIs:

- Query Web Services API
- Administration Web Services API

Crawler Plug-in API

Oracle SES includes an extensible crawler plug-in framework that lets you crawl and index proprietary document repositories with the Crawler Plug-in API.

Security APIs

Oracle SES also includes an extensible authentication and authorization framework. You use any identity management system to authorize users with the Identity Plug-in API, and you can define your own security model for each source with the Authorization Plug-in API.

URL Rewriter API

The URL Rewriter API is used by the crawler to filter and rewrite extracted URL links before they are inserted into the URL queue.

Query-time Authorization API

The Query-time Authorization API filters search results and access to document information at search time. Query-time filtering can be used in addition to, or instead of, ACLs.

Oracle Secure Enterprise Search Web Services APIs

Oracle SES includes the following Web Services APIs:

- **Query Web Services API**: Enables you to perform search queries; for example, search for "oracle benefits" and return all the documents. You can also customize the default Oracle SES ranking to create a more relevant search result list for your enterprise or configure clustering for customized applications.

- **Administration Web Services API**: Enables you to perform various administrative tasks; for example, start or stop a crawl schedule, check schedule status, get the estimated index fragmentation level, and perform index optimization.

  > **See Also:**
  >
  > - *Oracle Secure Enterprise Search Administration API Guide* for more information about Web services interface
  >
  > - *Oracle Secure Enterprise Search Java API Reference*
  >
  > - *Oracle SES Administration Tutorial* at
  >   http://www.oracle.com/pls/topic/lookup?ctx=E35215-01&id=tut_ses_administration

Oracle Secure Enterprise Search Web Services APIs let you write your own application to search and administer Oracle SES over the network. The APIs provide the following benefits:

- Applications can be deployed into any computer that connects to Oracle SES server through a standard Internet protocol.

- Web Services protocol is XML-based, which makes for easy application integration.

  > **Note:** The use of otext:: syntax is not in supported in a search string in the query application as well as in the query Web Services API.

Oracle SES also provides the client-side Java proxies for marshalling and parsing Web Services SOAP messages. Client applications can use the library instead of creating SOAP requests and parsing SOAP responses by themselves to access Oracle SES Web Services.

This section contains the following topics:

- Web Services APIs Installation

- Web Services Concepts

- Web Services Architecture

- Query Web Services Common Data Types

- Query Web Services Operations

- Query Web Services Query Syntax
- Query Web Services Example: Basic Search
- Query Web Services Example: Customizing Relevancy
- Client-Side Query Java Proxy Library

Web Services APIs Installation

Oracle SES Web Services runs on the Oracle SES middle tier standalone Oracle WebLogic Server. They are installed and configured as part of the default installation. You can use Oracle SES Web Services immediately. Follow the same middle tier administration steps to start and stop Oracle SES Web Services. Note that the Query Web Service client API should be run using JDK version 6 or later.

WebLogic provides a default Oracle SES Web Services administrator console. The administrator console URL is the same as the Oracle SES Web Services URL.

Query Web Services Location

The Query Web service is located at the following address for an Oracle SES installation:

```
http://host:port/search/query/OracleSearch
```

For example, if your Oracle SES middle tier is running on host `myhost` and the port number is `8888`, then the Query Web Services URL is the following:

```
http://myhost:8888/search/query/OracleSearch
```

You can obtain the following information from the administrator console:

- Oracle SES Query WSDL description
- List of Web Services messages and operations
- Client-side Java proxies and source codes

Administration Web Services Location

The Administration Web service is located at the following address for an Oracle SES installation:

```
http://host:port/search/api/admin/AdminService
```

You can obtain the following information from the administrator console:

- Oracle SES Administration WSDL description
- List of Web Services messages and operations
- Client-side JavaScript stub

Web Services Concepts

Oracle SES Web Services consists of a remote procedure call (RPC) interface to Oracle SES that enables the client application to invoke operations on Oracle SES over the network. The client application uses WSDL specification published by Oracle SES Web Services URL to send a request message using Simple Object Access Protocol (SOAP). The server then responds to the client application with a SOAP response message.

This section explains the following concepts:

- Web Services

- Simple Object Access Protocol
- Web Services Description Language

Web Services

A Web Service is a software application identified by a URI whose interfaces and binding are capable of being defined, described, and discovered by XML artifacts. A Web Service supports direct interactions with other software applications using XML-based messages and internet-based products.

A Web Service does the following:

- Exposes and describes itself: A Web Service defines its functionality and attributes so that other applications can understand it. By providing a WSDL file, a Web Service makes its functionality available to other applications.

- Allows other services to locate it on the Web: A Web Service can be registered in a UDDI registry so that applications can locate it.

- Can be invoked: After a Web Service has been located and examined, the remote application can invoke the service using an Internet standard protocol.

- Web Services are of either request and response or one-way style, and they can use either synchronous or asynchronous communication. However, the fundamental unit of exchange between Web Services clients and Web Services, of either style or type of communication, is a message.

Simple Object Access Protocol

The Simple Object Access Protocol (SOAP) is a lightweight XML-based protocol for exchanging information in a decentralized distributed environment. SOAP supports different styles of information exchange, including RPC-oriented and message-oriented exchange. RPC style information exchange allows for request-response processing, where an endpoint receives a procedure-oriented message and replies with a correlated response message. Message-oriented information exchange supports organizations and applications that must exchange messages or other types of documents where a message is sent, but the sender might not expect or wait for an immediate response. Message-oriented information exchange is also called document style exchange.

SOAP has the following features:

- Protocol independence
- Language independence
- Platform and operating system independence
- Support for SOAP XML messages incorporating attachments (using the multipart MIME structure)

Web Services Description Language

The Web Services Description Language (WSDL) is an XML format for describing network services containing RPC-oriented and message-oriented information. Programmers or automated development tools can create WSDL files to describe a service and can make the description available over the Internet. Client-side programmers and development tools can use published WSDL specifications to obtain information about available Web Services and to build and create proxies or program templates that access available services.

Web Services Architecture

Oracle SES Web Services is powered by the Oracle SES middle tier Oracle WebLogic Server. The implementation, configuration, and deployment of Oracle SES Web Services follow the procedures and standards provided by Oracle WebLogic Server.

Oracle SES WSDL defines the operations and messages for Oracle SES Web Services. The message exchange of Oracle SES Web Services is RPC style, in which the contents of the SOAP message body conform to a structure that specifies a procedure and includes a set of parameters or a response with a result and any additional parameters.

Oracle SES SOAP messages use HTTP binding where a SOAP message is embedded in the body of a HTTP request and a SOAP message is returned in the HTTP response.

The following diagram illustrates the architecture of Oracle SES Web Services:
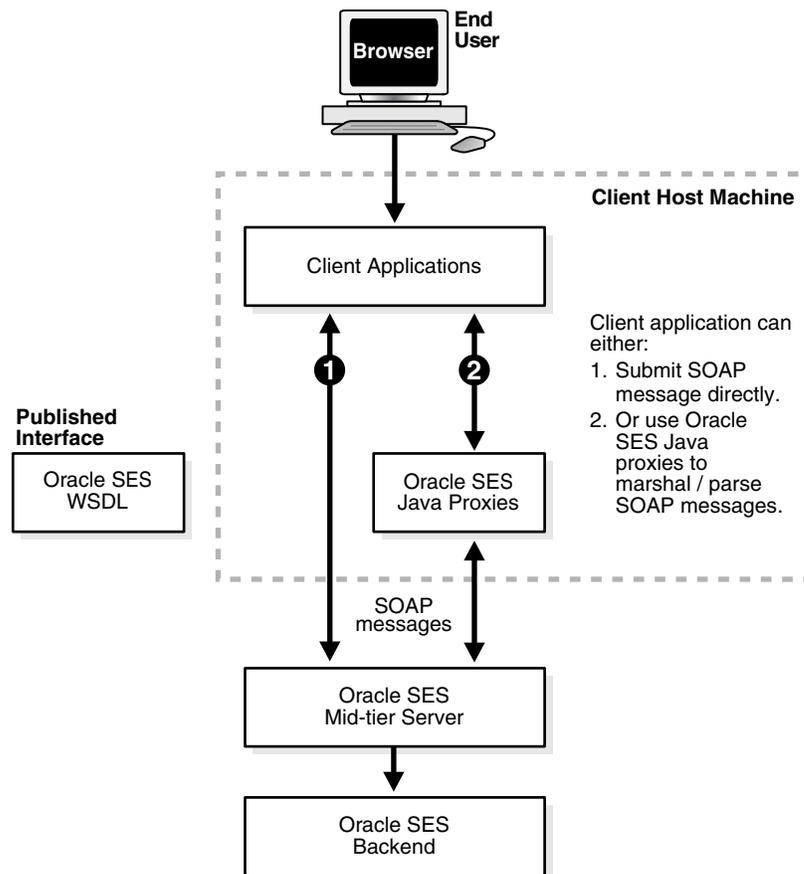
Development Platforms

You can implement client applications using platforms that support SOAP, such as Oracle JDeveloper, Microsoft .NET, or Apache Axis. These platforms allow you to automatically create code using the Oracle SES WSDL interface. Include the generated code along with the application logic to create a request, invoke the Web Services, and interpret the response.

Query Web Services Common Data Types

This section contains the following topics:

- Base Data Types

- XML-to-Java Data Type Mappings
- Complex Types
- Array Types

Base Data Types

Oracle Secure Enterprise Search Web Services use the following base data types:

Table 13–1   Base Data Types

Base Type	Description	Example
xsd:Boolean	Boolean	true, false
xsd:dateTime	Date	2005-12-31
xsd:int	Integer	256
xsd:long	Long integer	12345678900
xsd:string	String	Oracle Secure Enterprise Search

XML-to-Java Data Type Mappings

The mapping between XML schema data types and Java data types depends on the SOAP development environment. The following table shows mappings for the Oracle JDeveloper environment:

Table 13–2   XML-to-Java Type Mappings

XML Schema	Oracle JDeveloper
xsd:Boolean	java.lang.Boolean
xsd:dateTime	java.util.Date
xsd:int	java.lang.Integer
xsd:long	java.lang.Long
xsd:string	java.lang.String

Complex Types

Oracle SES Web Services uses the following complex data types:

- OracleSearchResult
- ResultElement
- SCElement
- DataGroup
- Attribute
- Filter
- Node
- AttributeLOVElement
- SessionContextElement
- Status
- Language

OracleSearchResult The search result container. It has the following elements:

- `returnCount`: A Boolean value indicating whether the result includes the count estimate for the hit list.

- `estimatedHitCount`: The estimated count of the search result; -1 means the search result does not return the estimated hit count.

- `dupRemoved`: A Boolean value indicating whether **near duplicate documents** have been removed from search result.

- `dupMarked`: A Boolean value indicating whether **near duplicate documents** have been marked in the search result. If `dupRemoved` is true, then `dupMarked` is always false.

- `resultElements`: An array of `resultElement`, which represents the actual hit list.

- `suggestedLinks`: An array of `suggestedLink` for the given search.

- `query`: The actual search string, which uses Oracle SES query syntax.

- `altKeywords`: Alternate keywords (suggestions) for a given search. If you set switch `ses.qapp.multiple_alternate_keywords` to `true`, then multiple alternate keywords are returned for a search term. The following terms are returned:

  - **ALTERNATE_SPELLING**: Contains alternate words from the dictionary.

  - **EXPANDED_QUERY**: Contains keywords from the alternate keywords page when Auto-Expand is selected. These keywords are automatically included in the query that is sent to Oracle Text.

  - **ALTERNATE_QUERY**: Contains keywords from the alternate keywords page when Auto-Expand is not selected.

- `startIndex`: The start index of search results.

- `docsReturned`: The number of search matches returned.

ResultElement This is the data type for search result element. It has the following elements:

- `author`: Primary author of the document

- `description`: Description of the document

- `url`: URL of the document

- `snippet`: Keywords in context (KWIC) of the document

- `title`: Title of the document

- `lastModified`: Last modified date of the document

- `mimetype`: Mime type of the document

- `score`: Oracle Text score of the document

- `docID`: Document ID

- `language`: Language of the document

- `contentLength`: Content length of the document

- `signature`: Signature of the document

- `infoSourceID`: InfoSource ID of the document

- `infoSourcePath`: InfoSource path of the document

- `groups`: Array of groups to which the document belongs

- `isDuplicate`: Boolean value indicating whether this document is a near duplicate of another document in the result list

- `hasDuplicate`: Boolean value indicating whether this document has one or more near duplicates in the result list

- `fedID`: Federated instance ID, used to track which federated instance the document is fetched from

- `customAttributes`: Array of custom nondefault attributes extracted from/for the document during crawling

  `CustomAttribute` encapsulates the name and value of the custom (user-defined) attribute. The name of the attribute is represented by actual name and type of the attribute in name_type format. For example, the string attribute foo is represented as `foo_STRING`. All Date attributes use the format `mm/dd/yyyy`.

SCElement Suggested content from a provider. It has the following elements:

- `name`: name of the suggested content provider

- `content`: suggested content from the provider. The content is a byte array of the XML or HTML content

DataGroup The source group. It has the following elements:

- `groupID`: Source group ID

- `groupName`: Source group name

- `groupDisplayName`: Display name for the source group

Attribute The data type for search attribute. It has the following elements:

- `id`: Search attribute ID

- `name`: Internal name of search attribute

- `displayName`: Display name of search attribute

- `type`: The search attribute type. Value is either number, string, or date.

Filter The data type for filter condition (predicate). It has the following elements:

- `attributeId`: Search attribute ID

- `attributeType`: Search attribute type. Value is either number, string, or date.

- `operator`: Operator of the filter condition

  - If `attributeType` is string, then it should be either equals or contains.

  - If `attributeType` is number or date, then it should be either greaterthan, greaterthanequals, lessthan, lessthanequals, or equals.

- `attributeValue`: Value of the filter condition (predicate)

  - For string type attribute, the value is simply the string itself.

  - For number type attribute, the value should be represented by a string consisting of an optional sign, (+) or (-), followed by a sequence of zero or more decimal digits ("the integer"), optionally followed by a fraction. The fraction consists of a decimal point followed by zero or more decimal digits. The string must contain at least one digit in either the integer or the fraction.

- For date type attribute, the value should be in the format mm/dd/yyyy, where mm is the month (01~12), dd is the date (01~31), yyyy is the year (for example, 2005)

Examples:

- If the filter condition is Title contains Oracle Secure Enterprise Search, then the client application must look up the attribute ID of search attribute Title and include the following (element, value) pairs:

    - attributeID = 1 (assuming the search attribute id of Title is 1)

    - operator = contains

    - attributeValue = Oracle Secure Enterprise Search

- If the filter condition is Price greater than 1000, then the client application must look up the attribute ID of search attribute Price and include the following (element, value) pairs:

    - attributeID = 2 (assuming the search attribute id of 'Price' is 2)

    - operator = greaterthan

    - attributeValue = 1000

Node This is the data type for the **infosource** node. It has the following elements:

- id: Infosource node ID

- fedId: Federated instance ID, used to track which federated instance the node belongs to

- name: Name of the node

- docCount: Number of documents under the node. If the value is -1, then there exists documents under the node but the count cannot be shown.

- hasChildren: Indicates if the node has any children

- fullpath: Full path of the category node

- fullpathIds: The IDs of each node in the full path

AttributeLOVElement This is the element of AttributeLOV, the list of search attribute values. It has the following elements:

- value: Attribute value (internal value)

- displayValue: Display value

SessionContextElement This data structure is used to store authentication information for the search user in the form of a name-value pair, which can be used during query-time authorization filtering of the results. It has the following elements:

- name: Name of the authentication attribute

- value: Value of the authentication attribute

Status This is the status of the request. It has the following elements:

- status: Status code. Value is either successful or 'ailed

- message: Status message. Value is null, or an error message if the status is 'ailed

Language This is the language data type. It has the following elements:

- `languageName`: Name of the language

- `languageDisplayName`: Display name (translated name) of the language

Array Types

Oracle Secure Enterprise Search Web Services uses the following complex array types:

- `AttributeArray`: Array of `Attribute`

- `AttributeLOVElementArray`: Array of `AttributeLOVElement`

- `CustomAttributeArray`: Array of `CustomAttribute`

- `SCElementArray`: Array of `SCElement`

- `DataGroupArray`: Array of `DataGroup`

- `FilterArray`: Array of `Filter`

- `IntArray`: Array of `int`

- `LanguageArray`: Array of `Language`

- `NodeArray`: Array of `Node`

- `ResultElementArray`: Array of `ResultElement`

- `SessionContextElementArray`: Array of `SessionContextElement`

- `StringArray`: Array of `String`

Query Web Services Operations

This section contains the following topics:

- Overview of Query Web Services Operations

- Authentication Operations

- Search Operations

- Browse Operations

- Metadata Operations

- Search Hit Operations

- User Feedback Operations

Overview of Query Web Services Operations

Oracle Secure Enterprise Search provides the following categories of Web Services operations:

- **Authentication**: Authenticates a user's access to Oracle SES. The operation is only required if the user performs secure search.

- **Search**: Runs a search on Oracle SES and obtains a hit list along with information such as estimated hit count, **near duplicate documents** in the result list, suggested links, and alternate keywords for the search. Gets suggested content from external providers for the given query. You can also customize the default Oracle SES ranking to create a more relevant search result list for your enterprise or configure clustering for customized Oracle SES applications.

- **Metadata**: Obtains the search metadata, such as the list of source groups, the list of supported languages, or the list of search attributes.

- **Search Hit**: Obtains the search result details, such as the cached version of search result and in-links and out-links of the search hit.

- **User Feedback**: Sends user feedback to Oracle SES, such as user-submitted URLs.

> **See Also:** "Query Web Services Operations" on page 13-10

Authentication Operations

This section describes the following authentication operations:

- loginRequest Message

- loginResponse Message

- logoutRequest Message

- logoutResponse Message

- setSessionContextRequest Message

- setSessionContextResponse Message

- proxyLoginRequest Message

- proxyLoginResponse Message

loginRequest Message Requests Oracle SES to authenticate the search user. It consists of the following parameters:

- username: User name for the search user. This value is *not* case-sensitive.

- password: Password for the search user

```
<message name="loginRequest">
    <part name="username"      type="xsd:string"/>
    <part name="password"      type="xsd:string"/>
</message>
```

loginResponse Message Contains the return status for the loginRequest message.

```
<message name="loginResponse">
    <part name="return"      type="typens:Status"/>
</message>
```

logoutRequest Message Used when the user logs out from the search application.

```
<message name="logoutRequest">
</message>
```

logoutResponse Message Contains the return status for the logoutRequest message.

```
<message name="logoutResponse">
    <part name="return"      type="typens:Status"/>
</message>
```

setSessionContextRequest Message Passes authentication information for the search user, which can be used during query-time filtering. It consists of the following parameter:

- sessionContext: An array of SessionContextElement. This array stores the authentication information needed for the query-time authentication filtering in the form of name-value pairs.

```
<message name="setSessionContextRequest">
    <part name="sessionContext"      type="typens:SessionContextElementArray"/>
</message>
```

> **Note:** Login and logout Web Services calls cause Oracle SES to automatically set or reset the AUTH_USER value in the session context that is passed to the query-time filter. This session context attribute cannot be overwritten explicitly through the setSessionContext call.

setSessionContextResponse Message Contains the return status for the setSessionContext message.

```
<message name="setSessionContextResponse">
  <part name="return"              type="typens:Status"/>
</message>
```

proxyLoginRequest Message Logs in the end user to Oracle SES using proxy authentication. It consists of following parameters:

- username: User name of the proxy user

- password: Password of the proxy user

- searchUser: User name of the end user

```
<message name="proxyLoginRequest">
  <part name="username"            type="xsd:string"/>
  <part name="password"            type="xsd:string"/>
  <part name="searchUser"          type="xsd:string"/>
</message>
```

The proxy user must be a federation trusted entity created on the Oracle SES instance.

> **See Also:** "Federation Trusted Entities" on page 6-14

proxyLoginResponse Message This message contains the return status for the proxyLoginRequest message.

```
<message name="proxyLoginResponse">
  <part name="return"          type="typens:Status"/>
</message>
```

Search Operations

This section describes the following search operations:

- doOracleAdvancedSearch Message

- doOracleFetchSearch Message

- doOracleOrganizedSearch Message

- doOracleSearch Message

- doOracleSearchResponse Message

- doOracleBrowseSearch Message

- doOracleBrowseSearchResponse Message

- doOracleSimpleSearch Message

- doOracleSimpleSearchResponse Message

- doOracleFacetSearch Message

- getSuggestedContent Message

- getSuggestedContentResponse Message

doOracleAdvancedSearch Message Invokes Oracle SES advanced search and returns search results. It consists of the following parameters:

- `query`: The search string. This should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 13-29 for details.

- `startIndex`: Index of the first document in the hit list to be returned. The default is 1 if not set explicitly.

- `docsRequested`: The maximum number of documents in the hit list to be returned. The default is 10 if not set explicitly.

- `dupRemoved`: Boolean flag to enable or disable duplicate removal. If turned on, then duplicate documents in the hit list are removed. The default is false if not set explicitly. Note: The `dupMarked` switch has no effect when `dupRemoved` is turned on.

- `dupMarked`: Boolean flag to enable or disable duplicate detection. The default is false if not set explicitly. Note: The `dupMarked` switch has no effect when `dupRemoved` is turned on.

- `groups`: Data source groups that the search is restricted to. The default is all groups if not set explicitly.

- `queryLang`: Language of the query. This is equivalent to Locale. The default is English (`en`) if not set explicitly. This is used in relevancy boosting.

- `docLang`: Language of the documents to restrict the search. If the value is not set explicitly, then search is performed against documents of all the languages.

- `returnCount`: Boolean flag to fetch the total hit count with the result. The default is false if not set explicitly.

- `filterConnector`: Connector between all the filters: "and" indicates that the documents in the hit list must satisfy all the filters, and "or" indicates that the documents in the hit list must satisfy at least one filter. The default is "and" if not set explicitly.

- `filters`: An array of filters. Each filter is a restriction on the search result. Filters are connected by the `filterConnector`. The default is null (no filter applies to the search result) if not set explicitly.

- `fetchAttributes`: Array of integers representing the IDs of custom or nondefault attributes to be fetched with the search result

- `searchControls`: XML string to specify advanced filter conditions and ranking parameters

> **Note:** The attribute filter `LastModifiedDate` uses the format `mm/dd/yyyy`.

```
public OracleSearchResult doOracleAdvancedSearch(
      String query,
      Integer startIndex,
      Integer docsRequested,
      Boolean dupRemoved,
      Boolean dupMarked,
      DataGroup[] groups,
      String queryLang,
      String docLang,
```

```
        Boolean returnCount,
        String filterConnector,
        Filter[] filters,
        Integer[] fetchAttributes,
        String searchControls)
    throws Exception
```

doOracleFetchSearch Message Invokes Oracle SES fetch search and returns fetch results. It consists of the following parameters:

- `query`: The search string. This should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 13-29 for details.

- `targetDocIdList`: Target document ID list, most likely from a cluster node.

- `startIndex`: Index of the first document in the hit list to be returned. The default is 1 if not set explicitly.

- `docsRequested`: Maximum number of documents in the hit list to be returned. The default is 10 if not set explicitly.

- `queryLang`: Language of the query. This is equivalent to Locale. The default is English (en) if not set explicitly. This is used in relevancy boosting.

- `fetchAttributeNames`: Array of names of custom or nondefault attributes to be fetched with the search result.

- `groupAttr`: Attribute used for grouping.

- `sortAttrList`: List of sorting attribute settings.

- `clusterList`: List of cluster configurations.

```
public OracleResultContainer doOracleFetchSearch(
        String query,
        String[] targetDocIdList,
        Integer startIndex,
        Integer docsRequested,
        String queryLang,
        String[] fetchAttributeNames,
        GroupAttribute groupAttr,
        SortAttribute[] sortAttrList,
        ClusterConfig[] clusterList)
    throws Exception
```

doOracleOrganizedSearch Message This invokes Oracle SES organized search and returns search results. It consists of the following parameters:

`query`: The search string. This should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 13-29 for details.

`topN`: Top N search result for grouping, sorting, and clustering.

`startIndex`: Index of the first document in the hit list to be returned. The default is 1 if not set explicitly.

`docsRequested`: Maximum number of documents in the hit list to be returned. The default is 10 if not set explicitly.

`dupRemoved`: Boolean flag to enable or disable duplicate removal. If turned on, duplicate documents in the hit list are removed. The default is false if not set explicitly. Note: The `dupMarked` switch has no effect when `dupRemoved` is turned on.

`dupMarked`: Boolean flag to enable or disable duplicate detection. The default is false if not set explicitly. The `dupMarked` switch has no effect when `dupRemoved` is turned on.

`groups`: Data source groups that the search is restricted to. The default is all groups if not set explicitly.

`queryLang`: Language of the query. This is equivalent to Locale. The default is English (`en`) if not set explicitly. This is used in relevancy boosting.

`docLang`: Language of the documents to restrict the search. If the value is not set explicitly, then search is performed against documents of all the languages.

`returnCount`: Boolean flag to fetch the total hit count with the result. The default is false if not set explicitly.

`filterConnector`: Connector between all the filters: "and" indicates that the documents in the hit list must satisfy all filters, "or" indicates that the documents in the hit list must satisfy at least one filter. The default is "and" if not set explicitly.

`filters`: An array of filters. Each filter is a restriction on the search result. Filters are connected by the `filterConnector`. The default is null (no filter applies to the search result) if not set explicitly.

`fetchAttributeNames`: Array of names of custom or nondefault attributes to be fetched with the search result.

`searchControls`: XML string to specify advanced filter conditions and ranking parameters.

`groupAttr`: Attribute used for grouping.

`sortAttrList`: List of sorting attribute settings.

`clusterList`: List of cluster configurations.

```
public OracleResultContainer doOracleOrganizedSearch(
        String query,
        Integer topN,
        Integer startIndex,
        Integer docsRequested,
        Boolean dupRemoved,
        Boolean dupMarked,
        DataGroup[] groups,
        String queryLang,
        String docLang,
        Boolean returnCount,
        String filterConnector,
        Filter[] filters,
        String[] fetchAttributeNames,
        String searchControls,
        GroupAttribute groupAttr,
        SortAttribute[] sortAttrList,
        ClusterConfig[] clusterList)
    throws Exception
```

doOracleSearch Message This is the main message for the search application. It consists of the following parameters:

- `query`: A search string. It must be a valid string and it cannot be null. The search string should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 13-29 for details.

- `startIndex`: The index of the first result to be returned. For example, if there are 67 results, you might want to start at 20. The default is 1 if not set explicitly.

- `docsRequested`: The maximum number of results to be returned. The default is 10 if not set explicitly.

- `dupRemoved`: Enable or disable duplicate removal. If turned on, then the search result eliminate all **near duplicate documents** from the result list. The `dupMarked` switch has no effect when `dupRemoved` is turned on. The default is false if not set explicitly.

- `dupMarked`: Enable or disable duplicate detection. If `dupRemoved` is turned off and `dupMarked` is turned on, then the search result keeps all **near duplicate documents** from the result list and marks them as duplicates. If `dupRemoved` is turned on, then the `dupMarked` switch has no effect. The default is false if not set explicitly.

- `groups`: Limit the search result to the documents from specified source groups. The default is for all groups if not set explicitly.

- `queryLang`: The query language argument should be a valid ISO language code. These codes are the lower-case, two-letter codes as defined by ISO-639. Examples: "en" for English and "de" for German. The default is English ("en") if not set explicitly. This is used for relevancy boosting.

- `docLang`: Set the language of the documents to limit the search. If the value is not set explicitly, then search is performed against documents of all the languages.

- `returnCount`: Set to true to return total hit count with the result. The default is false if not set explicitly.

- `filterConnector`: The connector between all filters: "and" indicates the search result must satisfy all filters, "or" indicates the search result must satisfy at least one filter. The default is "and" if not set explicitly.

- `filters`: An array of filters. Each filter is a restriction on search results. Filters are connected by `filterConnector`. The default is null (no filter applies to the search result) if not set explicitly.

- `fetchAttributes`: Array of integers representing the nondefault attribute IDs to be fetched in the `resultElements`. The default is null (or set one int value '0'), so no attributes other than default-attributes are fetched in the `resultElements`.

```
<message name="doOracleSearch">
    <part name="query"            type="xsd:string"/>
    <part name="startIndex"       type="xsd:int"/>
    <part name="docsRequested"    type="xsd:int"/>
    <part name="dupRemoved"       type="xsd:boolean"/>
    <part name="dupMarked"        type="xsd:boolean"/>
    <part name="groups"           type="typens:DataGroupArray"/>
    <part name="queryLang"        type="xsd:string"/>
    <part name="docLang"          type="xsd:string"/>
    <part name="returnCount"      type="xsd:boolean"/>
    <part name="filterConnector"  type="xsd:string"/>
    <part name="filters"          type="typens:FilterArray"/>
    <part name="fetchAttributes"  type="typens:IntArray"/>
</message>
```

doOracleSearchResponse Message This message returns the search result in `OracleSearchResult` data type.

```
<message name="doOracleSearchResponse">
    <part name="return"      type="typens:OracleSearchResult"/>
</message>
```

doOracleBrowseSearch Message This message restricts a search to a particular node. It consists of the following parameters:

- `query`: A search string. It must be a valid string, and it cannot be null. The search string should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 13-29 for more details.

- `nodeID`: The ID of the node to restrict the search to.

- `fedID`: The ID of the federated instance the parent node belongs to (null for local node).

- `startIndex`: The index of the first result to be returned. For example, if there are 67 results, then you might want to start at 20. The default is 1 if not set explicitly.

- `docsRequested`: The maximum number of results to be returned. The default is 10 if not set explicitly.

- `dupRemoved`: Enable or disable duplicate removal. If turned on, then the search result eliminate all **near duplicate documents** from the result list, and the `dupMarked` switch have no effect when `dupRemoved` is turned on. The default is false if not set explicitly.

- `dupMarked`: Enable or disable duplicate detection. If `dupRemoved` is turned off and `dupMarked` is turned on, then the search result keeps all **near duplicate documents** from the result list and marks them as duplicates. If `dupRemoved` is turned on, then the `dupMarked` switch has no effect. The default is false if not set explicitly.

- `queryLang`: The query language argument should be a valid ISO language code. These codes are the lower-case, two-letter codes as defined by ISO-639. Examples: "en" for English and "de" for German. The default is English ("en") if not set explicitly. This is used for relevancy boosting.

- `docLang`: Set the language of the documents to limit the search. If the value is not set explicitly, then search is performed against documents of all the languages.

- `returnCount`: Set to true to return total hit count with the result. The default is false if not set explicitly.

- `fetchAttributes`: Array of integers representing the nondefault attribute IDs to be fetched in the `resultElements`. The default is null (or set one int value '0'), so no attributes other than default-attributes are fetched in the `resultElements`.

```
<message name="doOracleBrowseSearch">
  <part name="query"            type="xsd:string"/>
  <part name="nodeID"           type="xsd:string"/>
  <part name="fedID"            type="xsd:string"/>
  <part name="startIndex"       type="xsd:int"/>
  <part name="docsRequested"    type="xsd:int"/>
  <part name="dupRemoved"       type="xsd:boolean"/>
  <part name="dupMarked"        type="xsd:boolean"/>
  <part name="queryLang"        type="xsd:string"/>
  <part name="docLang"          type="xsd:string"/>
  <part name="returnCount"      type="xsd:boolean"/>
  <part name="fetchAttributes"  type="typens:IntArray"/>
</message>
```

doOracleBrowseSearchResponse Message Returns the search result in `OracleSearchResult` data type.

```
<message name="doOracleBrowseSearchResponse">
    <part name="return"      type="typens:OracleSearchResult"/>
</message>
```

doOracleSimpleSearch Message A simplified form of the doOracleSearch message. In this message you do not need to specify the advanced search parameters that are specified in the doOracleSearch message. It consists of following parameters:

- `query`: A search string. It must be a valid string and it cannot be null. The search string should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 13-29 for details.

- `startIndex`: The index of the first result to be returned. For example, if there are 67 results, you might want to start at 20. The default is 1, if not set explicitly.

- `docsRequested`: The maximum number of results to be returned. The default is 10, if not set explicitly.

- `dupRemoved`: Enable or disable duplicate removal. If turned on, then the search result eliminates all **near duplicate documents** from the result list. The `dupMarked` switch has no effect when `dupRemoved` is turned on. The default is false if not set explicitly.

- `dupMarked`: Enable or disable duplicate detection. If `dupRemoved` is turned off and `dupMarked` is turned on, then the search result keeps all **near duplicate documents** from the result list and marks them as duplicates. If `dupRemoved` is turned on, then the `dupMarked` switch has no effect. The default is false if not set explicitly.

- `returnCount`: Set to true to return total hit count with the result. The default is false if not set explicitly.

```
<message name="doOracleSimpleSearch">
  <part name="query"            type="xsd:string"/>
  <part name="startIndex"       type="xsd:int"/>
  <part name="docsRequested"    type="xsd:int"/>
  <part name="dupRemoved"       type="xsd:boolean"/>
  <part name="dupMarked"        type="xsd:boolean"/>
  <part name="returnCount"      type="xsd:boolean"/>
</message>
```

doOracleSimpleSearchResponse Message Returns the search result in `OracleSearchResult` data type.

```
<message name="doOracleSimpleSearchResponse">
  <part name="return"     type="typens:OracleSearchResult"/>
</message>
```

doOracleFacetSearch Message Invokes Oracle SES faceted search and returns search results. It consists of the following parameters:

- `query`: The search string. This should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 13-29 for details.

- `topN`: Top N search result for grouping, sorting, and clustering.

- `startIndex`: Index of the first document in the hit list to be returned. The default is 1 if not set explicitly.

- `docsRequested`: The maximum number of documents in the hit list to be returned. The default is 10 if not set explicitly.

- `dupRemoved`: Boolean flag to enable or disable duplicate removal. If turned on, then duplicate documents in the hit list are removed. The default is false if not set explicitly. Note: The `dupMarked` switch has no effect when `dupRemoved` is turned on.

- `dupMarked`: Boolean flag to enable or disable duplicate detection. The default is false if not set explicitly. Note: The `dupMarked` switch has no effect when `dupRemoved` is turned on.

- `groups`: Data source groups that the search is restricted to. The default is all groups if not set explicitly.

- `queryLang`: Language of the query. This is equivalent to Locale. The default is English (`en`) if not set explicitly. This is used in relevancy boosting.

- `docLang`: Language of the documents to restrict the search. If the value is not set explicitly, then search is performed against documents of all the languages.

- `returnCount`: Boolean flag to fetch the total hit count with the result. The default is false if not set explicitly.

- `filterConnector`: Connector between all the filters: "and" indicates that the documents in the hit list must satisfy all the filters, and "or" indicates that the documents in the hit list must satisfy at least one filter. The default is "and" if not set explicitly.

- `filters`: An array of filters. Each filter is a restriction on the search result. Filters are connected by the `filterConnector`. The default is null (no filter applies to the search result) if not set explicitly.

- `fetchAttributeNames`: Array of strings representing the names of custom or nondefault attributes to be fetched with the search result.

- `groupAttr`: Attributes used for grouping.

- `sortAttrList`: List of sortable attributes settings.

- `clusterList`: List of cluster configurations.

- `facetPaths`: List of facet values in the form of facet path string. Each facet path denotes one facet tree, with the root being the facet name. For example, `Bath/Faucets/Shower` searches for `/Faucets/Shower` facet values under the `Bath` facet.

> **Note:** If `facetPaths` parameter is set to `null` or is assigned only the root level facet names, then the facets are used for grouping the search results and are not used for filtering the search results. The search results will contain all the documents matching the query criteria and will be grouped according to the root level facets.

You can also specify per-facet limits, such as - `returnFacetDocCount`, `minFacetDocCount`, `maxFacetChildren`, and `sortBy` - as comma separated name-value pairs in the facet path string.

For example:

```
facetPaths[0] = Facet=Location, Path=Location/California, minFacetDocCount=1,
maxFacetChildren=5, returnFacetDocCount=true, sortBy=ALPHA_ASC

facetPaths[1] = Facet=Food, Path=Food/Vegetable/Cabbage, minFacetDocCount=1,
maxFacetChildren=5, returnFacetDocCount=true , sortBy= ALPHA_ASC

facetPaths[2] = Facet=Author, Path=Author, minFacetDocCount=1,
maxFacetChildren=10, returnFacetDocCount=true, sortBy=COUNT_DES
```

You can also specify per-request facet limits using separate parameters - `returnFacetDocCount`, `minFacetDocCount`, `maxFacetChildren`, and `sortBy` - of the `doOracleFacetSearch()` API method, which are described in the following paragraphs. Per-facet limits take precedence over per-request facet limits. If these facet limits are not specified at per-facet level as well as at per-request level, then default values are used for the facet limits.

> **Note:** If a facet path parameter value has either character "=" or character "," in it, then that value must be enclosed in double quotes. If a parameter value contains any double quote character, then the backslash character "\" should be used to escape the double quote character.
>
> For example:
>
> ```
> Facet="Price, \"random\""
> ```

- `returnFacetDocCount`: A boolean flag indicating whether to return document count for each facet. Default is `TRUE`.

- `minFacetDocCount`: Minimum number of document count to return for a facet node for each facet. Default is 0.

- `maxFacetChildren`: Maximum number of child nodes to return for each facet and facet node. Default is 5. The 0 value denotes no child nodes should be returned. The -1 value denotes all the child nodes should be returned.

- `sortBy`: A constant value denoting the sort order of facet child nodes. The supported values are:

  - `COUNT_ASC`: Sort facet nodes in the ascending order of document count.

  - `COUNT_DES`: Sort facet nodes in the descending order of document count.

  - `ALPHA_ASC`: Sort facet nodes in the alphabetically ascending order of their names.

  - `ALPHA_DES`: Sort facet nodes in the alphabetically descending order of their names.

  - `FIXED`: Return facet nodes in the order they are configured in the facet tree.

  Default is `COUNT_DES`.

- `searchControls`: An XML string to specify advanced filter conditions and ranking parameters.

```
public OracleFacetResultContainer doOracleFacetSearch(
        String query,
        Integer topN,
        Integer startIndex,
        Integer docsRequested,
        Boolean dupRemoved,
        Boolean dupMarked,
        DataGroup[] groups,
        String queryLang,
        String docLang,
        Boolean returnCount,
        String filterConnector,
        Filter[] filters,
        String[] fetchAttributeNames,
        GroupAttribute groupAttr,
```

```
        SortAttribute[] sortAttrList,
        ClusterConfig clusterList,
        String[] facetPaths,
        Boolean returnFacetDocCount,
        Integer minFacetDocCount,
        Integer maxFacetChildren,
        String sortBy,
        String searchControls)
   throws Exception
```

getSuggestedContent Message Returns the suggested content for the given query. It consists of the following parameters:

- `query`: Query string

- `returnType`: Format in which the content is to be returned, either `html` or `xml`. If no style sheet is configured for a given provider, then the return type is the return type of the content returned by the provider, regardless of whether `html` or `xml` is specified.

```
<message name="getSuggestedContent">
  <part name="query"                    type="xsd:string"/>
  <part name="returnType"               type="xsd:string"/>
</message>
```

getSuggestedContentResponse Message Returns the suggested content for the query.

```
<message name="getSuggestedContentResponse">
  <part name="return"  type="typens:SCElementArray"/>
</message>
```

Browse Operations

This section describes the following browse operations:

- getInfoSourceNodesRequest Message

- getInfoSourceNodesResponse Message

- getInfoSourceAncestorNodesRequest Message

- getInfoSourceAncestorNodesResponse Message

- getInfoSourceNodeRequest Message

- getInfoSourceNodeResponse Message

getInfoSourceNodesRequest Message Obtains the list of info source nodes given the parent node ID. It consists of the following parameters:

- `parentNodeID`: The node ID for which all children nodes are returned. If it is not set, then the message returns all the root nodes.

- `fedID`: The ID of the federated instance the parent node belongs to (null for local node).

- `locale`: A two letter representation of locale. The default is English ("en") if not set explicitly.

```
<message name="getInfoSourceNodesRequest">
    <part name="parentNodeID"     type="xsd:string"/>
    <part name="fedID"            type="xsd:string"/>
    <part name="locale"           type="xsd:string"/>
</message>
```

getInfoSourceNodesResponse Message Returns an array of info source nodes.

```
<message name="getInfoSourceNodesResponse">
    <part name="nodes"    type="typens:NodeArray"/>
</message>
```

getInfoSourceAncestorNodesRequest Message Obtains the full path of a node, from root to node, given an info source node. It consists of the following parameters:

- `nodeID`: The node ID for which all the nodes in the path from root to node are returned; nodeID must be set and it cannot be null.

- `locale`: A two letter representation of locale. The default is English ("en") if not set explicitly.

```
<message name="getInfoSourceAncestorNodesRequest">
    <part name="nodeID"        type="xsd:string"/>
    <part name="locale"        type="xsd:string"/>
</message>
```

> **Note:** The `getInfoSourceAncestorNode` messages have been deprecated in Oracle SES.

getInfoSourceAncestorNodesResponse Message Returns an array of info source ancestor nodes.

```
<message name="getInfoSourceAncestorNodesResponse">
    <part name="nodes"    type="typens:NodeArray"/>
</message>
```

getInfoSourceNodeRequest Message Retrieves a particular node. It consists of the following parameters:

- `nodeID`: The node ID of the node to get, nodeID must be set and it cannot be null.

- `fedID`: The ID of the federated instance the parent node belongs to (null for local node).

- `locale`: A two letter representation of Locale, the default is English ("en") if not set explicitly.

Message format:

```
<message name="getInfoSourceNodeRequest">
   <part name="nodeID        "type="xsd:string"/>
   <part name="fedID"        "type="xsd:string"/>
   <part name="locale        "type="xsd:string"/>
</message>
```

getInfoSourceNodeResponse Message This message returns the node requested.

```
<message name="getInfoSourceNodeResponse">
    <part name="node    "type="typens:Node"/>
</message>
```

Metadata Operations

This section describes the following metadata operations:

- getLanguageRequest Message

- getLanguageResponse Message

- getDataGroupsRequest Message
- getDataGroupsResponse Message
- getAttributesRequest Message
- getAttributesResponse Message
- getAllAttributesRequest Message
- getAllAttributesResponse Message
- getAttributeLOVRequest Message
- getAttributeLOVResponse Message
- getFacetNodes Message
- addTag Message
- deleteTag Message

getLanguageRequest Message Obtains all the languages supported by Oracle SES. It is used by the client application to display the list of languages. It consists of the following parameter:

`locale`: A two letter representation of locale. The default is English (`en`) if not set explicitly.

```
<message name="getLanguagesRequest">
    <part name="locale"        type="xsd:string"/>
</message>
```

getLanguageResponse Message

This message returns all supported languages.

```
<message name="getLanguagesResponse">
    <part name="return"        type="typens:LanguageArray"/>
</message>
```

getDataGroupsRequest Message Requests for all source groups defined in Oracle SES. It is used by the client application to show all source groups in the search page, such that the end user can restrict their search results within one or multiple source groups. It consists of the following parameter:

`locale`: A two letter representation of locale. The default is English (`en`) if not set explicitly.

```
<message name="getDataGroupsRequest">
    <part name="locale"        type="xsd:string"/>
</message>
```

getDataGroupsResponse Message Returns all source groups defined in Oracle SES.

```
<message name="getDataGroupsResponse">
    <part name="groups"        type="typens:DataGroupArray"/>
</message>
```

getAttributesRequest Message Obtains a list of search attributes that applied to the given source groups. It consists of the following parameters:

- `locale`: A two letter representation of locale. The default is English (`en`) if not set explicitly.

- `groups`: Limit the request to the attributes from specified source groups. The default is all groups if not set explicitly.

- `groupConnector`: The connector between all groups: "and" indicates the response is the attributes available in the set of source groups by finding the intersection of each group's attributes, "or" indicates the response is the attributes available in the set of source groups by finding the union of each group's attributes. The default is "or" if not set explicitly.

```
<message name="getAttributesRequest">
    <part name="locale"         type="xsd:string"/>
    <part name="groups"         type="typens:DataGroupArray"/>
    <part name="groupConnector"  type="xsd:string"/>
</message>
```

getAttributesResponse Message Returns an array of search attributes.

```
<message name="getAttributesResponse">
    <part name="return"       type="typens:AttributeArray"/>
</message>
```

getAllAttributesRequest Message Obtains all search attributes defined in Oracle SES. It consists of the following parameter:

`locale`: A two letter representation of locale. The default is English (`en`) if not set explicitly.

```
<message name="getAllAttributesRequest">
  <part name="locale" type="xsd:string"/>
</message>
```

getAllAttributesResponse Message Returns all search attributes defined in Oracle SES.

```
<message name="getAllAttributesResponse">
    <part name="return"       type="typens:AttributeArray"/>
</message>
```

getAttributeLOVRequest Message Obtains the LOV items given a search attribute. It consists of the following parameters:

- `attribute`: A search attribute for the LOV (list of values) requested.

- `locale`: A two letter representation of locale. The default is English ("en") if not set explicitly.

```
<message name="getAttributeLOVRequest">
  <part name="attribute"    type="typens:Attribute"/>
  <part name="locale"       type="xsd:string"/>
</message>
```

getAttributeLOVResponse Message Returns an array of search attribute LOV elements.

```
<message name="getAttributeLOVResponse">
  <part name="return"    type="typens:AttributeLOVElementArray"/>
</message>
```

getFacetNodes Message Returns an array of facet child nodes having non-zero document count for facet tree paths. It consists of the following parameters:

- `facetPaths`: An array of facet tree path for which child facet nodes are to be returned.

You can also specify per-facet limits, such as - `returnFacetDocCount`, `minFacetDocCount`, `maxFacetChildren`, and `sortBy` - as comma separated name-value pairs in the facet path string.

For example:

```
facetPaths[0] = Facet=Location, Path=Location/California, minFacetDocCount=1,
maxFacetChildren=5, returnFacetDocCount=true, sortBy=ALPHA_ASC

facetPaths[1] = Facet=Food, Path=Food/Vegetable/Cabbage, minFacetDocCount=1,
maxFacetChildren=5, returnFacetDocCount=true , sortBy= ALPHA_ASC

facetPaths[2] = Facet=Author, Path=Author, minFacetDocCount=1,
maxFacetChildren=10, returnFacetDocCount=true, sortBy=COUNT_DES
```

You can also specify per-request facet limits using separate parameters - `returnFacetDocCount`, `minFacetDocCount`, `maxFacetChildren`, and `sortBy` - of the `getFacetNodes()` API method, which are described in the following paragraphs. Per-facet limits take precedence over per-request facet limits. If these facet limits are not specified at per-facet level as well as at per-request level, then default values are used for the facet limits.

> **Note:** If a facet path parameter value has either character "=" or character "," in it, then that value must be enclosed in double quotes. If a parameter value contains any double quote character, then the backslash character "\" should be used to escape the double quote character.
>
> For example:
>
> ```
> Facet="Price, \"random\""
> ```

- `returnFacetDocCount`: A boolean flag indicating whether to return document count for each facet. Default is `TRUE`.

- `minFacetDocCount`: Minimum number of document count to return for a facet node in each facet. Default is 0.

- `maxFacetChildren`: Maximum number of child nodes to return for each facet. Default is 5. The 0 value denotes no child nodes should be returned. The -1 value denotes all the child nodes should be returned.

- `sortBy`: A constant value denoting the sort order of facet child nodes. The supported values are:

  - `COUNT_ASC`: Sort facet nodes in the ascending order of document count.

  - `COUNT_DES`: Sort facet nodes in the descending order of document count.

  - `ALPHA_ASC`: Sort facet nodes in the alphabetically ascending order of their names.

  - `ALPHA_DES`: Sort facet nodes in the alphabetically descending order of their names.

  - `FIXED`: Return facet nodes in the order they are configured in the facet tree.

  Default is `COUNT_DES`.

- `locale`: A two letter representation of locale. The default is English (`en`) if not set explicitly.

- searchControls: An XML string to specify advanced filter conditions and ranking parameters.

```
public FacetNode[] getFacetNodes(
        String[] facetPaths,
        Boolean returnFacetDocCount,
        Integer minFacetDocCount,
        Integer maxFacetChildren,
        String sortBy,
        String locale,
        String searchControls)
    throws Exception
```

addTag Message Specifies a tag for a document. Returns boolean value of `true` if the tag addition operation is successful, else returns `false`. It consists of the following parameters:

- docID: Document ID.

- fedID: Federation ID.

- tag: Name for the tag.

- username: Name of the user adding the tag.

```
public boolean addTag(
        Integer docID,
        Integer fedID,
        String tag,
        String username)
    throws Exception
```

deleteTag Message Deletes a tag for a document. Returns boolean value of `true` if the tag deletion operation is successful, else returns `false`. It consists of the following parameters:

- docID: Document ID.

- fedID: Federation ID.

- tag: Name for the tag.

- username: Name of the user deleting the tag.

- ctlParams: Control parameters.

```
public boolean deleteTag(
        Integer docID,
        Integer fedID,
        String tag,
        String username,
        String ctlParams)
    throws Exception
```

Search Hit Operations

This section describes the following search hit operations:

- getCachedPageRequest Message

- getCachedPageResponse Message

- getInLinksRequest Message

- getInLinksResponse Message

- getOutLinksRequest Message
- getOutLinksResponse Message
- logUserClickRequest Message
- logUserClickResponse Message

getCachedPageRequest Message Obtains the cached version of a document given the document ID and the search string. The search string is highlighted in the output. It consists of the following parameters:

- `query`: The search string.
- `docID`: The document ID to be fetched.
- `fedID`: The federated instance ID, used to track which federated instance the document is fetched from.

```
<message name="getCachedPageRequest">
    <part name="query"          type="xsd:string"/>
    <part name="docID"          type="xsd:int"/>
    <part name="fedID"          type="xsd:string"/>
</message>
```

getCachedPageResponse Message Returns the byte array of the cached HTML page.

```
<message name="getCachedPageResponse">
    <part name="return"          type="xsd:base64Binary"/>
</message>
```

getInLinksRequest Message Obtains all the incoming links for a given search hit (document). It consists of the following parameters:

- `docID`: The document ID for the incoming links to be fetched. It must be a valid document ID and it cannot be null.
- `maxNum`: The maximum number of incoming links requested. The default is 25 if not set explicitly.
- `fedID`: The federated instance ID, used to track which federated instance the document is fetched from.

```
<message name="getInLinksRequest">
    <part name="docID"                    type="xsd:int"/>
    <part name="maxNum"                   type="xsd:int"/>
    <part name="fedID"                    type="xsd:string"/>
</message>
```

getInLinksResponse Message Returns an array of incoming link URL strings.

```
<message name="getInLinksResponse">
    <part name="return"      type="typens:StringArray"/>
</message>
```

getOutLinksRequest Message Obtains all the outgoing links for a given search hit (document). It consists of the following parameters:

- `docID`: The document ID for the outgoing links to be fetched. It must be a valid document ID and it cannot be null.
- `maxNum`: The maximum number of outgoing links requested. The default is 25 if not set explicitly.

- `fedID`: The federated instance ID, used to track which federated instance the document is fetched from.

```
<message name="getOutLinksRequest">
    <part name="docID"                type="xsd:int"/>
    <part name="maxNum"               type="xsd:int"/>
    <part name="fedID"                type="xsd:string"/>
</message>
```

getOutLinksResponse Message This message returns an array of outgoing link URL strings.

```
<message name="getOutLinksResponse">
    <part name="return"       type="typens:StringArray"/>
</message>
```

logUserClickRequest Message This message logs the user's click. It consists of the following parameters:

- `queryID`: ID of the submitted search.

- `urlID`: ID of the document that the user clicked.

- `infosourceID`: Infosource ID. If none, then -1 is used as the default value

- `position`: The position of the document in the result list. For example, the first hit on the page or ninth hit on the page.

- `fedID`: Federation ID. Specifies the federated instance on which the document resides.

```
<message name="logUserClickRequest">
    <part name="queryID"       type="xsd:int"/>
    <part name="urlID"         type="xsd:int"/>
    <part name="infoSourceID"  type="xsd:int"/>
    <part name="position"      type="xsd:int"/>
    <part name="fedID"         type="xsd:string"/>
</message>
```

logUserClickResponse Message Returns the URL of the clicked document.

```
<message name="logUserClickResponse">
  <part name="url"               type="xsd:string"/>
</message>
```

User Feedback Operations

This section describes the following user feedback operations:

- submitUrlRequest Message

- submitUrlResponse Message

submitUrlRequest Message Submits a URL to Oracle SES so that it crawls and indexes the URL. This operation consists of the following parameter:

- `url`: The URL to be submitted to the crawler so it can be crawled next time. It must be a valid URL and it cannot be null.

```
<message name="submitUrlRequest">
     <part name="url"        type="xsd:string"/>
</message>
```

submitUrlResponse Message Returns the status, which consists of two strings. The first is the submission status, which is either `successful` or `failed`. The second string is the error message when the submission status is `failed`.

```
<message name="submitUrlResponse">
    <part name="return"     type="typens:Status"/>
</message>
```

Query Web Services Query Syntax

This section describes the query syntax used in the Oracle Secure Enterprise Search Search API.

Search Term

A search term can be a single word, a phrase, or a special search term. For example, if the search string is `oracle secure enterprise search`, then there are four search terms in the search string: `oracle`, `secure`, `enterprise`, and `search`. If the search string is `oracle "secure enterprise search"`, then there are two search terms in the search string: `oracle` and `"secure enterprise search"`.

Search terms are case insensitive so that different cases are treated the same. For example, searching `oracle`, `Oracle`, or `ORACLE` returns the same search result.

Phrase

A phrase is a string enclosed in double-quotes ("). It can contain one or multiple words.

Operators

The following operators are defined in the query syntax:

- Plus [+]: The plus operator specifies that the search term immediately following it must be found in all matching documents. For example, searching for [Oracle +Applications] only finds documents that contain the word "Oracle" and "Applications". In a multiple word search, you can attach a [+] in front of every token including the very first token. You can also attach a [+] in front of a phrase enclosed in double-quotes ("). But there should be no space between the [+] and the search term.

- Minus [-]: The minus operator specifies that the search term immediately following it cannot appear in any document included in the search result. For example, searching for [Oracle -Applications] only finds documents that do not contain the word "Applications". In a multiple word search, you can attach a [-] in front of every token except the very first token. It can be a single word or a phrase, but there should be no space between the [-] and the token.

- Asterisk [*]: The asterisk specifies a wildcard search. For example, searching for the string [Ora*] finds documents that contain all words beginning with "Ora" such as "Oracle" and "Orator". You can also insert an asterisk in the middle of a word. For example, searching for the string [A*e] finds documents that contain words such as "Apple" or "Ape".

Default Search: Implicit AND

By default, Oracle SES searches all of your search terms and relevant variations of the terms you entered. You do not need to include any operators (like `AND`) between terms. The order of the terms in the search affects the search results.

Word Separator

Use one or more spaces to separate each of the search terms.

Filter Conditions (Advanced Conditions)

Oracle SES query syntax only supports 'Site' and 'File type' filter conditions. It does not support any other filter conditions (advanced conditions) such as title, author, or last modified date. To restrict your search with other filter conditions, you can specify them in the Web Services API message `doOracleSearch`.

Special Search Terms

Oracle SES supports the use of several special search terms that allow the user or search administrator to access additional capabilities of the Oracle SES. Following is the list of special search terms:

Exclude Search Term You can exclude a word from your search by putting a minus sign [-] immediately in front of the term you want to exclude from the search results. Exclusion does not work with stop words.

Example: `oracle -search`

Negative search is not allowed unless there is another positive search term. For example:

`-search` is an invalid search.

`oracle -search` is a valid search.

Wildcard Search You can use an asterisk to match any number of characters in the middle or the end of a search term. You cannot place it at the beginning, such as searching for *earch.

Example: `Ora*`

Phrase Search Search for complete phrases by enclosing them in quotation marks. Words marked in this way appear together in all results exactly as entered.

Example: `"oracle secure enterprise search"`

Site Restricted Search If you know the specific Web site you want to search, but are not sure where the information is located within that site, then search only within the specific Web site. Enter the search followed by the string `site:` followed by the host name.

Example: `oracle site:example.com`

Notes:

- Domain restriction is not supported, because Oracle SES does not support left-truncated wildcard search (such as *.example.com)

- The exclusion operator (-) can be applied to this search term to remove a Web site from consideration in the search.

- Site restricted search term is implicit AND with other search terms.

- Only one site restriction is allowed. Also, you cannot have both site inclusion and exclusion in the search string. For example, the following search string is invalid:

  `oracle search  site:www.oracle.com  -site:otn.oracle.com`

File Type Restricted Search The search prefix `filetype:` filters the results returned to include only documents with the extension specified immediately after. There can be no space between `filetype:` and the specified extension.

Example: `oracle filetype:doc`

Notes:

- The exclusion operator (-) can be applied to this search term to remove a file type from consideration in the search.

- Only one file type can be included. The following extensions are supported: doc, htm, html, xml, ps, pdf, txt, rtf, ppt, and xls. doc, html, pdf, txt, rtf, ppt, xls.

- File type restricted search term is implicit AND with other search terms.

- Only one file type restriction is allowed. Also, you cannot have both file type inclusion and exclusion in the search string. For example, the following search string is invalid:

  ```
  oracle search  filetype:doc -filetype:pdf
  ```

Query Web Services Example: Basic Search

Following is a simple JSP application using Oracle Secure Enterprise Search proxy Java library to provide the basic search functionality:

```
<%@page contentType="text/html; charset=utf-8" %>
<%@page import = "java.util.Vector" %>
<%@page import = "java.net.URL" %>
<%@page import = "java.util.Properties" %>
<%@page import = "java.util.HashMap" %>
import javax.xml.rpc.Stub;
<%@page import = "oracle.search.query.webservice.client.*" %>

<%
  //
  // Get the search term entered by the user
  //
  String searchTerm = request.getParameter("searchTerm");
  if (searchTerm == null)  searchTerm = "";

  //
  // Define the result element array.
  //
  // ResultElement is a proxy Java class
  ResultElement[] resElemArray = null;
  int estimatedHitCount = 0;

  if (searchTerm != null && !"".equals(searchTerm))
  {
    //
    // Create the Oracle SES Web Services client stub
    //
    OracleSearchService stub = new OracleSearchService();

    //
    // Set the Oracle SES Web Services URL.
    // The URL is http://<host>:<port>/search/query/OracleSearch
    //
    stub.setSoapURL("http://staca19:7777/search/query/OracleSearch");
```

```
            //
            // Get the search result by calling OracleSearchService.doOracleSearch()
            //
            OracleSearchResult result = stub.doOracleSearch(searchTerm,
                                        new Integer(1),
                                        new Integer(10),
                                        Boolean.TRUE,
                                        Boolean.TRUE,
                                        null,
                                        "en",
                                        "en",
                                        Boolean.TRUE,
                                        null,
                                        null,
                                        null);
        //
        // Get the estimated hit count by calling
        estimatedHitCount = result.getEstimatedHitCount().intValue();

        // Get the search results
        resElemArray = result.getResultElements();
    }
%>

<HTML>
<HEAD>
    <TITLE>Oracle SES Web Services Demo </TITLE>
</HEAD>
<BODY>
<FORM name="searchBox" method="post" action="./DemoWS.jsp">
 <INPUT id="inputMain" type="text" size="40" name="searchTerm"
value="<%=searchTerm%>">
 <INPUT type="hidden"  name="searchTerm" value="<%= searchTerm %>">
 <INPUT type="submit" name="action" value="Search">
</FORM>
<BR><BR><BR>

<%
  //
  // Render the search results
  //
  if (resElemArray == null || resElemArray.length == 0)
  {
%>
  <H3> There are no matches for the search term </H3>
<%
  }
  else
  {
%>
  <H3> There are about <%=estimatedHitCount%> matches </H3>
<%
    for (int i=0; i<resElemArray.length; i++)
    {
      String title = resElemArray[i].getTitle();
      if (title == null) title = "Untitled Document";
%>
  <P>
    <B><A HREF="<%=resElemArray[i].getUrl()%>"><%=title%></A> </B>
    <BR>
```

```
    <%=resElemArray[i].getSnippet()%>
    <BR>
  </P>
<%
    }
  }
%>
</BODY>
</HTML>
```

Default-Factor Element

The default-factor element assigns a weight to an attribute.

```
<default-factor>
    <name>title</name>
    <weight>VERY HIGH</weight>
</default-factor>
```

Default factor attribute names are case-insensitive.

When a `default-factor` does not appear in the ranking XML string, Oracle SES takes the default weight for this ranking factor, unless default factors are disabled by `enable-all-default-factor`.

Oracle SES supports the following values for weight element: `empty` (Oracle SES uses the default weight), `none` (this attributes is not used in the ranking query), `very high`, `high`, `medium`, `low`, and `very low`.

Table 13–3 lists the `default-factor` names and weights:

Table 13–3   Oracle SES Default Attributes and Weights

Attribute	Weight
Title	High
Description	Medium
Reftext	High
Keywords	Medium
Subject	Low
Author	Medium
H1headline	Low
H2headline	Very low
Url	Low
Urldepth	High
Language Match	High
Linkscore	High

> **Note:** When display URL is not an absolute URL, for example,
> `idcplg?IdcService=GET_`
> `FILE&dDocName=CNT104862&allowInterrupt=1&Rendition=web`, or the
> depth of the URL does not have any bearing on the relevance of the
> document, then it is recommended to set the weight of `urldepth` to
> `NONE`, so that Oracle SES will not consider the depth of the URL in the
> final ranking order of the search result.
>
> For example,
>
> ```
> <default-factor>
>   <name>urldepth</name>
>   <weight>NONE</weight>
> </default-factor>
> ```

Query Web Services Example: Customizing Relevancy

The following is the signature of the method for advanced search:

```
public OracleSearchResult doOracleAdvancedSearch (
        String query,
        Integer startIndex,
        Integer docsRequested,
        Boolean dupRemoved,
        Boolean dupMarked,
        DataGroup groups[],
        String queryLang,
        String docLang,
        Boolean returnCount,
        String filterConnector,
        Filter filters[],
        Integer[] fetchAttributes,
        String searchControls)  throws Exception
```

The `searchControls` parameter accepts a XML string, which include the `filter` and
`ranking` elements.

```
<searchControls>
    <filter>  ... </filter>
    <ranking> ... </ranking>
</searchControls>
```

Filter Element

Filters for attribute search are passed in the `filter` element. All the various AND and
OR conditions on the attributes are specified in the XML. For example:

```
<filter>
   <operator type="and">
   <operator type="or">
      <attributefilter name="xxx" type="string" operation="equals" value="ttt"/>
      <attributefilter name="yyy" type="number"
        operation="greaterthan" value="22"/>
...
   </operator>
...
      <attributefiler name="aaa" type="number" operation="equals" value="22"/>
...
   </operator>
</filter
```

If the parameter `searchControls` is null, then `filters` and `filterConnector` are used to create advanced search; otherwise, they are ignored.

Ranking Element

The ranking XML string is expressed as `ranking` element in `searchControls`. The following is an example of `ranking` element:

```
<ranking>
   <global-settings>
      <enable-all-default-factor>TRUE</enable-all-default-factor>
   </global-settings>
   <default-factor>
      <!--default ranking factor -- >
      ...
   </default-factor>
   <default-factor>
      <!--default ranking factor -- >
      ...
   </default-factor>
   <custom-factor>
      <!--default ranking factor -- >
      ...
   </custom-factor>
   <custom-factor>
      <!--default ranking factor -- >
      ...
   </custom-factor>
   <sortConditions>
      <sortCondition order="ascending" type="system">Title</sortCondition>
   </sortConditions>
</ranking>
```

The following rules apply to the construction of ranking XML string:

- The whole ranking XML can be null, in which case default ranking is used.

- The ranking XML contains the elements `default-factor`, `custom-factor`, and `sortConditions`. All these elements can be null or absent at the same time.

- When `default-factor` is null or absent and when `custom-factor` is not null, default ranking is used with the effect of `custom-factor`.

- When `custom-factor` is null or absent, it does not have any impact on the ranking.

global-settings Element

The `global-settings` element contains parameter settings across ranking factors. The `ranking` element has an attribute called `enable-all-default-factor`, which accepts two values: `true` or `false`. (When this attribute is absent, `true` is taken as the default value.)

When `enable-all-default-factor` is `true`, all default attributes are included in ranking queries, unless some default attributes are explicitly excluded in `default-factor` elements.

When `enable-all-default-factor` is `false`, all default attributes are excluded in ranking queries, unless some default attributes are explicitly included in `default-factor` elements.

custom-factor Element

The custom-factor element lets you add more attributes for ranking. Any indexed search attribute can be a custom ranking attribute.

> **Note:** Adding custom attributes for relevancy ranking can downgrade search performance.

The custom-factor element has four elements: attribute-name, attribute-type, factor-type, and weight (or match depending on the factor-type).

```
<custom-factor>
    <attribute-name>author manager</attribute-name>
    <attribute-type>STRING</attribute-type>
    <factor-type>QUERY_FACTOR</factor-type>
    <weight>LOW</weight>
</custom-factor>
```

or

```
<custom-factor>
    <attribute-name>document quality</attribute-name>
    <attribute-type>STRING</attribute-type>
    <factor-type>STATIC_FACTOR<factor-type>
    <match>
       <value>good</value>
       <weight>HIGH</weight>
    </match>
    <match>
       <value>fair</value>
       <weight>MEDIUM</weight>
    </match>
    <match>
       <value>bad</value>
       <weight>VERY LOW</weight>
    </match>
</custom-factor>
```

or

```
<custom-factor>
    <attribute-name>sourcegroup</attribute-name>
    <attribute-type>SYSTEM</attribute-type>
    <factor-type>STATIC_FACTOR<factor-type>
    <match>
       <value>Corporate Site</value>
       <weight>VERY HIGH</weight>
    </match>
    <match>
       <value>Blogs</value>
       <weight>VERY LOW</weight>
    </match>
    <match>
       <value>Documentation</value>
       <weight>VERY LOW</weight>
    </match>
</custom-factor>
```

- The `attribute-name` values are literally matched against attribute name in Oracle SES. Any indexed search attribute name can be `attribute-name` value. The value of the `attribute-name` element is case-insensitive.

- The `attribute-type` element defines the type of the attribute. String and System attribute types are supported. Attribute-name and attribute-type in combination define a valid Oracle SES attribute.

> **Note:** System attribute type supports only the attribute-name *sourcegroup*, which represents the source group name. For System attribute type, the factor-type value must be "STATIC_FACTOR" as user's query is not applicable to source groups.

- For `factor-type`, Oracle SES supports two types of ranking for custom ranking attributes.

  - `QUERY_FACTOR`: The attribute value is matched against query terms. A positive match boosts the document based on specified weight. `QUERY_FACTOR` is a query-based ranking factor; for example, title and reftext. The `weight` element should appear for this custom ranking factor. For example, with the query "Roger Federer", if a document has a custom attribute publisher with the value "Roger Federer", then it could be relevant.

  - `STATIC_FACTOR`: Attribute value is matched against fixed values specified in the custom ranking factor. (The `match` element should appear for this custom ranking factor.) `STATIC_FACTOR` is not a query-based ranking factor. The fixed values specify qualities of the documents, such as the link score and the sources of documents. For example, assume that documents have been classified based on quality. Well-written documents are classified as good, and poorly-written documents are classified as bad. A good document should be ranked higher than a bad document, even though they are both matched against a query. You can specify in the API that a document having a good quality should be boosted in relevancy by a specified weight.

- The `match` element specifies the match values and corresponding match weights when the `factor-type` is `STATIC_FACTOR`. The following XML string is a example of `match` element:

```
<match>
   <value>bad</value>
   <weight>VERY LOW</weight>
</match>
```

- The `value` element is used to match the corresponding attribute value of this ranking factor. Only alphanumeric letters are allowed in the attribute value. The match is case-insensitive.

- The `weight` element has the identical syntax with `weight` element for default ranking element.

sortConditions Element

The `sortConditions` element defines one or more sort conditions for Absolute sort. Each Absolute sort condition is specified using the `sortCondition` child element of the `sortConditions` element. The value of the `sortCondition` element denotes the sortable attribute name.

The `order` parameter of the `sortCondition` element denotes the sort order for the specified sortable attribute. The sort order can have the value of either `ascending` or `descending`.

The `type` parameter of the `sortCondition` element denotes the attribute type of the specified sortable attribute. For a sortable *search* attribute, specify its value as `attribute`, and for a sortable *system* attribute, specify its value as `system`.

The following example specifies an Absolute sort criteria for sorting the search result based on the sortable system attribute `Title` in the ascending order.

```
<sortConditions>
    <sortCondition order="ascending" type="system">Title</sortCondition>
</sortConditions>
```

> **See Also:** "Configuring Sort Criteria using Sortable Attributes (Absolute Sort)" on page 5-22 for more information about configuring sort criteria for Absolute sort.

Applying Ranking Factors

The XML ranking text can be applied in two places:

- As a part of the `searchControls` element, the ranking factors can be used as an advanced control for each query execution through the Web services method. This is called **per-query ranking control**.

- The ranking factors specified in the `relevanceRanking` object of the Administration API are applied to all queries. This is called **instance-wide ranking control**.

In federated search, instance-wide ranking controls only applies to one instance. You must configure each instance for ranking customization separately.

If a conflict arises, the per-query ranking control specified in Web services method overrides the settings specified in instance-wide ranking control. That can include the following cases:

- Per-query and instance-wide ranking specify the same factor, the factor set by per-query is taken by Oracle SES.

- Instance-wide ranking control sets a ranking factor, but per-query ranking control does not mention. Oracle SES takes the factor set by instance-wide ranking control.

- Per-query ranking control sets a ranking factor, which instance-wide ranking controls does not mention. Oracle SES takes the factor set by per-query ranking control.

- If instance-wide ranking control sets `enable-all-default-factor` as false and per-query ranking control sets `enable-all-default-factor` as true, then Oracle SES takes the default attributes set explicitly by instance-wide ranking control plus the attributes set by per-query ranking controls, with the latter overriding the former.

Client-Side Query Java Proxy Library

Oracle SES also provides client-side Java proxies for marshalling and parsing Web Services SOAP messages. Client applications can use the library to access Oracle SES Web Services.

The proxy library includes the following Java classes, which are mapped to the corresponding Web Services data types and messages:

- `oracle.search.query.webservice.client.Attribute`
- `oracle.search.query.webservice.client.AttributeLOVElement`
- `oracle.search.query.webservice.client.ClusterAttribute`
- `oracle.search.query.webservice.client.ClusterConfig`
- `oracle.search.query.webservice.client.ClusterTree`
- `oracle.search.query.webservice.client.CustomAttribute`
- `oracle.search.query.webservice.client.DataGroup`
- `oracle.search.query.webservice.client.Filter`
- `oracle.search.query.webservice.client.GroupAttribute`
- `oracle.search.query.webservice.client.GroupingResult`
- `oracle.search.query.webservice.client.Language`
- `oracle.search.query.webservice.client.Node`
- `oracle.search.query.webservice.client.OracleSearchResult`
- `oracle.search.query.webservice.client.OracleSearchService`
- `oracle.search.query.webservice.client.ResultElement`
- `oracle.search.query.webservice.client.SCElement`
- `oracle.search.query.webservice.client.SessionContextElement`
- `oracle.search.query.webservice.client.SortAttribute`
- `oracle.search.query.webservice.client.Status`
- `oracle.search.query.webservice.client.SuggestedLink`

To compile and run your client application using the Oracle SES client-side Java proxy library, unzip the file *wls_home*/server/lib/wseeclient.zip, and include all the jar files present in this zip file in the Java CLASSPATH.

Also include the following jar files in the Java CLASSPATH:

- *wls_home*/server/lib/weblogic.jar
- *mw_home*/modules/org.apache.ant_1.7.1/lib/ant.jar
- *mw_home*/oracle_common/modules/oracle.webservices_11.1.1/oracle.webservices.standalone.client.jar
- *ses_home*/search/lib/search_client.jar
- *ses_home*/search/lib/plugins/cservices/jaxrpc.jar

Oracle Secure Enterprise Search Java SDK

The Oracle Secure Enterprise Search Java SDK contains the following APIs:

- Crawler Plug-in API
- Document Service API
- URL Rewriter API
- Security APIs
- Query-time Authorization API

> **See Also:** *Building Custom Crawlers* white paper for detailed
> information about how to build secure crawlers:
>
> http://www.oracle.com/technetwork/search/oses/overview/buidl
> ing-custom-crawlersjan12-07-1-133233.doc

Crawler Plug-in API

You can implement a crawler plug-in to crawl and index a proprietary document
repository. In Oracle SES, the proprietary repository is called a *user-defined source*. The
module that enables the crawler to access the source is called a crawler plug-in (or
connector).

The plug-in collects document URLs and associated metadata from the user-defined
source and returns the information to the Oracle SES crawler. The crawler starts
processing each URL as it is collected.

The crawler plug-in must be implemented in Java using the Oracle SES Crawler
Plug-in API. Crawler plug-ins can be either stored in the *ses_*
home/search/lib/plugins directory and accessed using their relative file path, or
stored in any other directory and accessed using their absolute file path.

These are the basic steps to build a crawler plug-in:

1. Compile and build the plug-in jar file.

   The Java source code for the plug-in first must be compiled into class files and put
   into a jar file in the *ses_home*/search/lib/plugins/ directory. The library needed
   for compilation is *ses_home*/search/lib/search_sdk.jar.

2. Create a source type.

   Before you can create a source for the crawler plug-in, you first must create a
   source type for it. From the Oracle SES Administration GUI, go to the **Global
   Settings - Source Types** page and provide the Java class name and jar file name
   (created in the previous step).

3. Create a source.

   From the **Home - Sources** page, create a source from the source type you just
   created. You also must define the parameter for the source type just created.

   For example, suppose you want to crawl /scratch/test on a Linux box for the file
   crawler plug-in. Specify the seed URL, for example,
   file://localhost/scratch/test.

4. Run the crawler plug-in.

   From the **Home - Schedules** page, start the schedule for the crawler.

This section includes the following topics:

- Crawler Plug-in Overview
- Crawler Plug-in Functionality

  > **See Also:** *Oracle SES Developer Tutorial for Building Custom Plug-Ins* at
  > http://www.oracle.com/pls/topic/lookup?ctx=E35215-01&id=tut_
  > ses_building_custom_plugins

Crawler Plug-in Overview

The following diagram illustrates the crawler plug-in architecture.

Crawler Plug-in API

Document content reader

Base Crawler

New Repository (for example, FileNet)

Crawler Plug-in (User implemented)

Built-in Source Types

Web Server

Database

Email

File System

OracleAS Portal
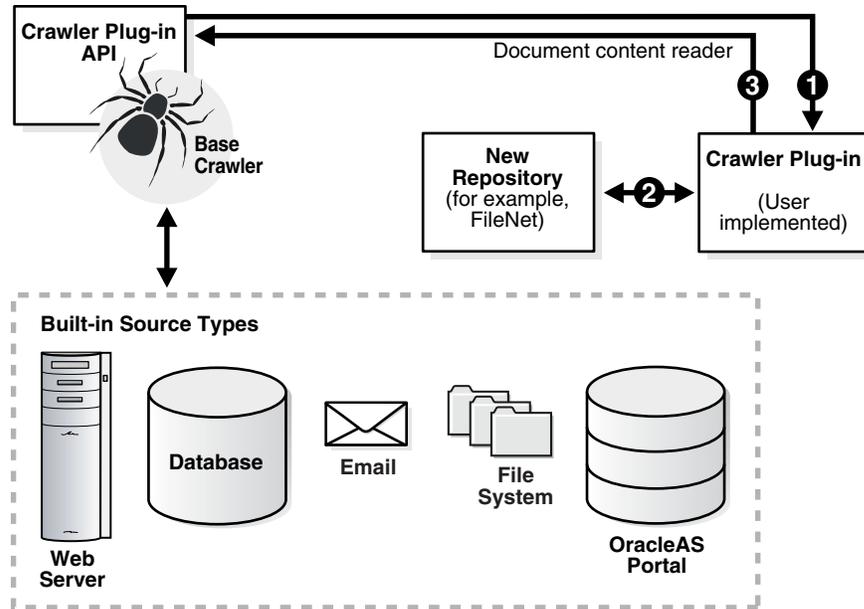
Two interfaces in the Crawler Plug-in API (`CrawlerPluginManager` and `CrawlerPlugin`) must be implemented to create a crawler plug-in. A crawler plug-in does the following:

- Provides the metadata of the document in the form of document attributes.

- Provides access control list information (ACL) if the document is protected.

- Maps each document attribute to a common attribute name used by end users.

- Optionally provides the list of URLs that have changed since a given time stamp.

- Optionally provides an access URL in addition to the display URL for the processing of the document.

- Provide the document contents in the form of a Java Reader. In other words, the plug-in is responsible for fetching the document.

- Can submit attribute-only documents to the crawler; that is, a document that has metadata but no document contents.

Document Attributes and Properties Document attributes, or metadata, describe document properties. Some attributes can be irrelevant to your application. The crawler plug-in creator must decide which document attributes should be extracted and saved. The plug-in also can be created such that the list of collected attributes are configurable. Oracle SES automatically registers attributes returned by the plug-in. The plug-in can decide which attributes to return for a document.

Library Path and Java Class Path Any other Java class needed by the plug-in should be included in the plug-in jar file. (You could add the paths for the additional jar files needed by the plug-in into the `Class-Path` of the `MANIFEST.MF` file in the plug-in jar file.) This is because Oracle SES automatically adds the plug-in jar file to the crawler Java class path, and Oracle SES does not let you add other class paths from the administration interface.

If the plug-in code also relies on a particular library file (for example, a `.dll` file on Windows or a `.so` file on UNIX), then the library must be put under the *ses_home*/lib directory or the *ses_home*/search/lib/plugins directory. The Java library path is set explicitly by the crawler to those locations.

You should use Java resource bundles instead of properties files whenever possible when developing a custom plug-in. If you must use the properties files as resource bundle files, then take these steps to ensure that the administration API loads the files properly.

To use properties files as resource bundles files:

1. Add the path of `sourceTypeJarPackageName` into the classpath variable in *ses_home*`/search/config/searchctl.conf`

2. Restart the middle tier.

Crawler Plug-in Restrictions The plug-in must handle mimetype rejection and large document rejection itself. For example, the plug-in should reject files it does not want to index based on its type or size, such as zip files. Also, plain text files, such as log files, can grow very large. Because the crawler reads HTML and plain text files into memory, it could run out of memory with very large files.

Crawler Plug-in Functionality

This section describes aspects of the crawler plug-in.

Source Registration Source registration is automated. After a source type is defined, any instance of that source type can be defined:

- Source name

- Description of the source; limit to 4000 bytes

- Source type ID

- Default language; default is `en` (English)

- Parameter values; for example:

```
seed - http://www.oracle.com
depth - 8
```

Source Attribute Registration You can add new attributes to Oracle SES by providing the attribute name and the attribute data type. The data type can be string, number, or date. Attributes returned by an plug-in are automatically registered if they have not been defined.

User-Implemented Crawler Plug-in The crawler plug-in has the following requirements:

- The plug-in must be implemented in Java.

- The plug-in must support the Java plug-in APIs defined by Oracle SES.

- The plug-in must return the URL attributes and properties.

- The plug-in must decide which document attributes Oracle SES should keep. Any attribute not defined in Oracle SES is registered automatically.

- The plug-in can map attributes to source properties. For example, if an attribute ID is the unique ID of a document, then the plug-in should return (document_key, 4) where `ID` has been mapped to the property `document_key` and its value is 4 for this particular document.

- If the attribute LOV is available, then the plug-in returns them upon request.

Crawler Plug-in APIs and Classes The Crawler Plug-in API is a collection of classes and interfaces of the Java SDK package `oracle.search.sdk.crawler` that is used to implement a crawler plug-in.

Table 13–4   Crawler Plug-in Interfaces and Classes

Interface/Class	Description
CrawlerPlugin	This interface is used by the crawler plug-in to integrate with the Oracle SES crawler.
	The Oracle SES crawler loads the plug-in manager class and invokes the plug-in manager API to obtain the crawler plug-in instance. Each plug-in instance is run in a thread execution.
CrawlerPluginManager	This interface is used to generate the crawler plug-in instances.
	It provides general plug-in information for automatic plug-in registration on the administration page for defining user-defined source types. It has the control on which plug-in object (if multiple implementations are available) to return in `getCrawlerPlugin` call and how many instances of the plug-in to return. If only one instance is returned, then the plug-in implementation must handle multi-threading execution.
	The `CrawlingThreadService` object pass in is thread-specific as the invocation of each `getCrawlerPlugin` call is initiated by each thread.
CrawlingThreadService	This interface is used by a crawler plug-in to perform crawl-related tasks. It has execution context specific to the crawling thread that invokes the plug-in `crawl()` method.
DataSourceService	This interface is implemented by the Oracle SES crawler and made available to the plug-in through the `GeneralService` object.
	This interface is used by a crawler plug-in to manage the current crawled document set.
DocumentAcl	This interface is used by a crawler plug-in to submit access control list (ACL) information for the document.
DocumentContainer	This interface is used by a crawler plug-in to submit or retrieve document information.
DocumentMetadata	This interface holds a document's attributes and properties for processing and indexing.
	This interface is used by a crawler plug-in to submit URL-related data to the crawler.
DocumentService	This interface is used by a document service plug-in to submit document attributes and/or document contents to the crawler.
DocumentServiceManager	This interface is used to register the document service plug-in. It is also used by the crawler to create a DocumentService object.
GeneralService	This interface provides Oracle SES service and implemented interface objects to the plug-in. It is implemented by the Oracle SES crawler and made available through plug-in manager initialization.
	This interface is used by a crawler plug-in to obtain Oracle SES interface objects.

Table 13–4   (Cont.) Crawler Plug-in Interfaces and Classes

Interface/Class	Description
Logger	This interface is used by a crawler plug-in to output messages to the crawler log file.
LovInfo	This interface is used by both the crawler and a source plug-in to set and retrieve attribute list of values (LOV) from the source.
ParameterValues	This interface for a crawler plug-in reads the value of the source parameter.
QueueService	This interface is implemented by the Oracle SES crawler and made available to the plug-in through the GeneralService object.
	This interface is used by the crawler plug-in to submit URL-related data to the crawler.
ParameterInfo	This class describes the general properties of a parameter.
DocumentServiceException	This provides a severity code to direct the crawler's response.
PluginException	This class encapsulates information about a data source plug-in-specific error.
ProcessingException	This class encapsulates information about errors from processing plug-in requests.
UserDefinedSecurityModel	The crawler manager class must implement this interface of the oracle.search.sdk.security package to use a crawler that uses attribute-based security.

Document Service API

The Document Service API is a part of the Crawler Plug-in API and is used to create a document service plug-in. A *document service plug-in* is a Java class that implements the document service API interface DocumentServiceManager of the oracle.search.sdk.crawler package and is registered using Oracle SES Administration GUI.

> **See Also:**   "Overview of Document Service" on page 4-14

To create and use a document service plug-in:

1. Create a Java file (for example, DocumentSummarizer.java) implementing the oracle.search.sdk.crawler.DocumentService interface methods init, close, and process.

2. Create another Java file (for example, DocumentSummarizerManager.java) implementing the oracle.search.sdk.crawler.DocumentServiceManager interface.

3. Compile all of the related Java files into class files. For example:

    ```
    $java_home/bin/javac -classpath ses_home/search/lib/search_sdk.jar
    DocumentSummarizer.java DocumentSummarizerManager.java
    ```

4. Package all the class files into a jar file under the *ses_home*/search/lib/plugins/doc directory. For example:

    ```
    $java_home/bin/jar cv0f ses_home/search/lib/plugins/doc/extractor/extractor.jar
    DocumentSummarizer.class DocumentSummarizerManager.class
    ```

The above document service plug-in jar file can be deployed either under the *ses_home*/search/lib/plugins/doc directory and accessed using its relative file path, or under any other directory and accessed using its absolute file path.

5. On the **Global Setting - Document Service** page, register the jar file as a new document service plug-in by providing the class name as oracle.search.plugin.doc.extractor.DocumentSummarizerManager, and providing either the relative jar file path of extractor/extractor.jar or the absolute jar file path (in case it is not deployed under the *ses_home*/search/lib/plugins/doc directory).

After registering a document service plug-in, you can create an instance for it as described in the section "Creating a Document Service" on page 4-15.

URL Rewriter API

A URL rewriter is a user supplied Java module that implements the Oracle SES Java SDK API interface UrlRewriter of the oracle.search.sdk.crawler package. When activated, it is used by the crawler to filter and rewrite extracted URL links before they are inserted into the URL queue.

The URL Rewriter API is included as part of the Crawler Plug-in SDK. The URL Rewriter API is used for Web sources.

Web crawling generally consists of the following steps:

1. Get the next URL from the URL queue. (Web crawling stops when the queue is empty.)

2. Fetch the contents of the URL.

3. Extract URL links from the contents.

4. Insert the links into the URL queue.

The generated new URL link is subject to all existing boundary rules.

There are two possible operations that can be done on the extracted URL link:

- Filtering: removes the unwanted URL link

- Rewriting: transforms the URL link

URL Link Filtering

Users control what type of URL links are allowed to be inserted into the queue with the following mechanisms supported by the Oracle SES crawler:

- robots.txt file on the target Web site; for example, disallow URLs from the /cgi directory

- Hosts inclusion and exclusion rules; for example, only allow URLs from www.example.com

- File path inclusion and exclusion rules; for example, only allow URLs under the /archive directory

- Mimetype inclusion rules; for example, only allow HTML and PDF files

- Robots metatag NOFOLLOW; for example, do not extract any link from that page

- Blacklist URLs; for example, URL explicitly singled out not to be crawled

With these mechanisms, only URL links that meet the filtering criteria are processed. However, there are other criteria that users might want to use to filter URL links. For example:

- Allow URLs with certain file name extensions

- Allow URLs only from a particular port number

- Disallow any PDF file from a particular directory

The possible criteria could be very large, so it is delegated to a user-implemented module that can be used by the crawler when evaluating an extracted URL link.

URL Link Rewriting

For some applications, due to security reasons, the URL crawled is different from the one seen by the end user. For example, crawling occurs on an internal Web site behind a firewall without security checking, but when queried by an end user, a corresponding mirror URL outside the firewall must be used.

A *display URL* is a URL string used for search result display. This is the URL used when users click the search result link. An *access URL* is a URL string used by the crawler for crawling and indexing. An access URL is optional. If it does not exist, then the crawler uses the display URL for crawling and indexing. If it does exist, then it is used by the crawler instead of the display URL for crawling.

For regular Web crawling, only display URLs are available. But in some situations, the crawler needs an access URL for crawling the internal site while keeping a display URL for external use. For every internal URL, there is an external mirrored URL.

For example:

```
http://www.example-qa.us.com:9393/index.html
http://www.example.com/index.html
```

When the URL link `http://www.example-qa.us.com:9393/index.html` is extracted and before it is inserted into the queue, the crawler generates a new display URL and a new access URL for it:

Access URL:

```
http://www.example-qa.us.com:9393/index.html
```

Display URL:

```
http://www.example.com/index.html
```

The extracted URL link is rewritten, and the crawler crawls the internal Web site without exposing it to the end user.

Another example is when the links that the crawler picks up are generated dynamically and can be different (depending on referencing page or other factor) even though they all point to the same page. For example:

```
http://compete3.example.com/rt/rt.wwv_media.show?p_type=text&p_id=4424&p_
currcornerid=281&p_textid=4423&p_language=us
```

```
http://compete3.example.com/rt/rt.wwv_media.show?p_type=text&p_id=4424&p_
currcornerid=498&p_textid=4423&p_language=us
```

Because the crawler detects different URLs with the same contents only when there is enough duplication, the URL queue could grow to a huge number of URLs, causing excessive URL link generation. In this situation, allow "normalization" of the extracted links so that URLs pointing to the same page have the same URL. The algorithm for

rewriting these URLs is application dependent and cannot be handled by the crawler in a generic way.

When a URL link goes through a rewriter, the following outcomes are possible:

- The link is inserted with no changes.

- The link is discarded; it is not inserted.

- A new display URL is returned, replacing the URL link for insertion.

- A display URL and an access URL are returned. The display URL might or might not be identical to the URL link.

> **Note:** URL rewriting is available for Web sources only.

To create and use a URL rewriter:

1. Create a new Java file implementing the `UrlRewriter` interface methods `open`, `close`, and `rewrite`.

2. Compile the rewriter Java file into a class file. For example:

   ```
   $java_home/bin/javac -classpath ses_home/search/lib/search_sdk.jar
   SampleRewriter.java
   ```

3. Package the rewriter class file into a jar file and place it either under the *ses_home*/search/lib/plugins/ directory, or under any other directory. For example:

   ```
   $java_home/bin/jar cv0f ses_home/search/lib/plugins/sample.jar
   SampleRewriter.class
   ```

4. Enable the `UrlRewriter` option and specify the rewriter class name (for example, `SampleRewriter`) and jar file path (for example, `sample.jar`, if the jar file is placed under the *ses_home*/search/lib/plugins/ directory. If you place the jar file in any other directory, then you must provide its absolute file path) in the Oracle SES Administration GUI **Home - Sources - Crawling Parameters** page of an existing Web source.

5. Crawl the target Web source by launching the corresponding schedule. The crawler log file confirms the use of the URL rewriter with the message *Loading URL rewriter "SampleRewriter"...*

   > **See Also:** *Oracle Secure Enterprise Search Java API Reference* for the Oracle SES Java SDK URL Rewriter API (`oracle.search.sdk.crawler` package)

Security APIs

In addition to the extensible crawler plug-in framework that lets you crawl and index proprietary document repositories (Crawler Plug-in API), Oracle SES also includes an extensible authentication and authorization framework. This lets you use any identity management system to authorize users (Identity Plug-in API). You can also define your own security model for each source (Authorization Plug-in API).

Oracle SES Java SDK provides the `oracle.search.sdk.security` package that contains the classes and interfaces related to the security API.

Identity Plug-in API

The Identity Plug-in API communicates with the identity management system to authenticate a user at login with a user name and password. It also provides a list of groups (or roles) for a specified user.

The identity plug-in manager manages initialization parameters and returns the `IdentityPlugin` object.

To add an identity plug-in, click **Register New Identity Plug-in** on the **Global Settings - Identity Management Setup** page, and enter the class name and jar file name for the identity plug-in manager.

Authorization Plug-in API

For sources with authorization requirements that do not fit the user/group model, an authorization plug-in provides a more flexible security model. (Authentication is still handled by an identity plug-in.)

With an authorization plug-in, a crawler plug-in can add security attributes similar to document attributes. The authorization plug-in is invoked at login time to build security filters onto the query string. The security filters are applied against the values of the security attributes for each document. Only documents whose security attribute values match the security filter are returned to the user. (All security attributes have string values.)

The authorization plug-in contains the following component:

- `ResultFilterPlugin`: This is an interface for the query-time authorization (QTA). When building the hit list, Oracle SES calls a result filter plug-in to check if the user is authorized to view each document. Only documents the user is authorized to view are listed in the hit list. The result filter can be used as the only security device., or it can be used with other security. The result filter can also be used to modify the title or display URL.

User-Defined Security Model

With the user-defined security model, Oracle SES displays an Authorization page before a new user-defined source can be completed. The `UserDefinedSecurityModel` interface provides a method that returns the name of the class implementing the `AuthorizationManager` interface and the names and types (`GRANT` or `DENY`) of the security attributes used to build the security filter for a given user.

If you must change the `AuthorizationManager` plug-in class name or jar file name, then you must turn off security for that source to allow the change. After changing and applying the ACL setting to **No Access Control List**, you can edit the `AuthorizationManager` details. The new `AuthorizationManager` should share the same security attribute model as the previous one.

> **Caution:** While security is turned off, any user can access the documents in the affected source.

> **See Also:** *Oracle Secure Enterprise Search Java API Reference* for the Oracle SES Java SDK Security API (`oracle.search.sdk.security` package)

Query-time Authorization API

Query-time authorization enables you to associate a Java class with a source that, at search time, validates every document fetched out of the Oracle SES repository belonging to the protected source. This result filter class can dynamically check access rights to ensure that the current search user has the credentials to view each document.

You can apply this authorization model to any source other than self service or federated sources. Besides acting as the sole provider of access control for a source, it can also be used as a post-filter. For example, a source can be stamped with a more generic ACL, while query-time authorization can be used to fine tune the results.

Overview of Query-time Authorization

Query-time authorization has the following characteristics:

- It allows dynamic access control at search time compared to more static ACL stamping.

- It filters documents returned to a search user.

- It controls the Browse functionality to determine whether a folder is visible to a search user.

- Optionally, it allows pruning of an entire source from the results to reduce performance costs of filtering each document individually.

- It is applicable to all source types except self service and federated sources.

- The result filter can modify the Title or Display URL for the result returned to the search user.

Query-time filtering is handled by class implementations of the `ResultFilterPlugin` interface of the `oracle.search.sdk.security` package.

Filtering Document Access

Filtering document access is handled by the `filterDocuments` method of the `ResultFilterPlugin` interface. The most common situation for filtering occur with a search request, in which this method is invoked with batches of documents from the result list. Based on the values returned by this method, all, some, or none of the documents might be removed from the results returned to the search user.

Access of individual documents is also controlled. For example, viewing a cached copy of a document or accessing the in-links and out-links requires a call into `filterDocuments` to determine the authorization for the search user.

Filtering Folder Browsing

The `ResultFilterPlugin` implementation is also responsible for controlling the access to, and visibility of folders in, the Browse application. If a folder belongs to a source protected by a query-time filter, then the folder name in the **Browse** page does not have a document count listed next to it. Instead, the folder shows a **view_all** link.

For performance reasons, it can be costly to determine the exact number of documents visible to the current search user for every query-time filtered folder displayed on a Browse page. This calculation requires that every document in every folder be processed by the filter. To prevent this comprehensive and potentially time-consuming operation, document counts are not used. Instead, folder visibility is explicitly determined by the query-time filter.

Based on the results from the `filterBrowseFolders` method, a folder might be hidden or shown in the Browse page. This result also controls access to the single folder browsing page, which displays the documents contained in a folder.

If the security of folder names is not a concern for a particular source, then the `filterBrowseFolders` method can blindly authorize all folders to be visible in the Browse application. After a folder is selected, the document list is still filtered through the `filterDocuments` method. This strategy should not be employed if folder names could reveal sensitive information.

If security is very critical, then it might be easiest to hide all folders for browsing. The documents from the source are still available for search queries from the Basic and Advanced Search boxes, but users are not able to browse the source in the **Browse** pages of the search application.

Limitations of folder filtering:

- The `filterBrowseFolders` method does not implicitly restrict access to subfolders. For example, if folder `/Miscellaneous/www.example.com/private` is hidden for a search user, then it is still possible for that user to view any subfolder, such as `/Miscellaneous/www.example.com/private/a/b`, if that subfolder is not also explicitly filtered out by this method. It would be possible to view this subfolder if the user followed a bookmark or outside link directly to the authorized subfolder in the Browse application.

- This method does not affect functionality outside of the Browse application. This is not a generic folder pruning method. Search queries and document retrieval outside of the Browse application are only affected by the `filterDocuments` and `pruneSource` methods.

Pruning Access to an Entire Source

The `ResultFilterPlugin` interface provides the ability to determine access privileges at the source level. This is achieved through calls to the `pruneSource` method. This method can be called in situations where there are a large number of documents or folders to be filtered. Authorizing or unauthorizing the entire source for a given user could provide a large performance gain over filtering each document individually.

The implementation of `ResultFilterPlugin` must not rely on this method to secure access to documents or folders. This method is strictly an optimization feature. There is no guarantee that it is invoked for any particular search request or document access. For example, when performing authorization for a single document, Oracle SES may call the `filterDocuments` method directly without invoking this method at all. Therefore, the `filterDocuments` and `filterBrowseFolders` methods must be implemented to provide full security without pruning.

Determining the Authenticated User

A query-time filter is free to define a search user's access privileges to sources and documents based on any criteria available. For example, a filter could be written to deny access to a source depending on the time of day.

In most cases, however, a filter imposes restrictions based on the authenticated user for that search request. The Oracle SES authenticated user name for a request is contained in the `oracle.search.sdk.query.RequestInfo` object. The steps for accessing this user name value depend on whether the request originated from the JSP search application or the Oracle SES Query Web Services interface. For either type of request, the key used to access the authenticated user name is the string value `AUTH_USER`. The user name is *not* case-sensitive.

This sample implementation of the `ResultFilterPlugin.getCurrentUserName` method illustrates how to retrieve the current authenticated user from either a JSP or Web Services request:

```
public String getCurrentUserName( RequestInfo req )
  throws PluginException
{
  HttpServletRequest servReq = req.getHttpRequest();
  Map sessCtx = req.getSessionContext();
  String user = null;

  if( servReq != null )
  {
    HttpSession session = servReq.getSession();
    if( session != null )
      user = ( String ) session.getAttribute( "AUTH_USER" );
  }

  else if( sessCtx != null )
  {
    // Web Service request
    user = ( String ) sessCtx.get( "AUTH_USER" );
  }

  if( user == null )
    user = "unknown";

  return user;
}
```

> **See Also:** "Authentication Methods" on page 11-12

Query-time Authorization Interfaces and Exceptions

The following Oracle SES Java SDK packages contain all the interfaces and exceptions for the Query-time Authorization API:

- `oracle.search.sdk.query`

- `oracle.search.sdk.common`

- `oracle.search.sdk.security`

To write a Query-time Authorization filter, implement the `oracle.search.sdk.security.ResultFilterPlugin` interface. The methods in this interface may throw instances of `oracle.search.sdk.common.PluginException`.

Objects that implement the `RequestInfo`, `DocumentInfo`, and `FolderInfo` interfaces of the `oracle.search.sdk.query` package are passed as arguments for filtering, but these interfaces do not need to be implemented by the filter writer.

The Query-time Authorization API contains the following interfaces and exceptions:

Table 13–5   Query-time Authorization Interfaces and Exceptions

Interface/Exception	Description
ResultFilterPlugin	This interface filters search results and access to document information at search time.
	If an object implementing this interface has been assigned to a source, then any search results or other retrieval of documents belonging to the source are passed through this filter before being presented to the end user.

Table 13–5 (Cont.) Query-time Authorization Interfaces and Exceptions

Interface/Exception	Description
PluginException	This exception is thrown by methods in the ResultFilterPlugin interface to indicate that a failure has occurred.
DocumentInfo	This interface represents information about a document that can be passed to a ResultFilterPlugin for filtering out documents.
FolderInfo	This interface represents information about a folder that can be passed to a ResultFilterPlugin to control folder browsing.
RequestInfo	This interface represents information about a request that can be passed to a ResultFilterPlugin for filtering out documents, folders, or entire sources.

> **See Also:** *Oracle Secure Enterprise Search Java API Reference* for the Oracle SES Java SDK Query-time Authorization API

Thread-Safety of the Filter Implementation

Classes that implement the ResultFilterPlugin interface should be designed to persist for the lifetime of a running search application. A single instance of ResultFilterPlugin generally handles multiple concurrent requests from different search end users. Therefore, the filterDocuments, pruneSource, filterBrowseFolders, and getCurrentUserName methods in this class must be both reentrant and thread-safe.

Compiling and Packaging the Query-Time Filter

To compile your query-time filter class, you must include at least the two following files in the Java CLASSPATH. These files can be found in the Oracle SES server directory.

- *ses_home*/search/lib/search_query.jar

- *ses_home*/lib/servlet.jar

Oracle recommends that you build a jar file containing your ResultFilterPlugin class (or classes) and any supporting Java classes. Place this jar file in a secure location for access by the Oracle SES server. If this jar file is compromised, then the security of document access in the search server can be compromised.

Your query-time filter might require other classes or jar files that are not included in the jar file you build and are not located in the Oracle SES class path. If so, add these files to the Class-Path attribute of the JAR file manifest. Include this manifest file in the jar file that you build.

If Oracle SES cannot locate a class used by a ResultFilterPlugin during run-time, then an error message is written to the log file and all documents from that source are filtered out for the search request being processed.

> **See Also:**
>
> http://docs.oracle.com/javase/6/docs/technotes/guides/jar/jar.html for more information about jar file manifests.

XML Connector Examples and Schemas

This appendix contains examples and schemas associated with the Oracle SES XML connector framework. This contains the following topics:

- Configuration File XML Schema Definition
- Control Feed Example
- Control Feed XML Schema Definition
- Data Feed Example
- Data Feed XML Schema Definition

> **See Also:** "Overview of XML Connector Framework" on page 4-11

Configuration File XML Schema Definition

The following example shows the XSD for the configuration file.

```
<?xml version="1.0" encoding="windows-1252"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns="http://xmlns.oracle.com/search/rsscrawlerconfig"
targetNamespace="http://xmlns.oracle.com/search/rsscrawlerconfig"
elementFormDefault="qualified">

<xsd:element name="rsscrawler">

   <xsd:annotation>
    <xsd:documentation>
     RSS crawler configuration paramters
    </xsd:documentation>
   </xsd:annotation>

<xsd:complexType>
 <xsd:sequence>

  <xsd:element name="sourceName" type="xsd:string" minOccurs="0"/>

  <xsd:element name="feedType" default="dataFeed">
   <xsd:simpleType>
    <xsd:restriction base="xsd:string">
     <xsd:enumeration value="controlFeed"/>
     <xsd:enumeration value="dataFeed"/>
     <xsd:enumeration value="directoryFeed"/>
    </xsd:restriction>
   </xsd:simpleType>
 </xsd:element>
```

```
<xsd:element name="feedLocation">
 <xsd:complexType>
  <xsd:simpleContent>
   <xsd:extension base="xsd:anyURI"/>
  </xsd:simpleContent>
 </xsd:complexType>
</xsd:element>

<xsd:element name="errorFileLocation" type="xsd:string" minOccurs="0">
 <xsd:annotation>
  <xsd:documentation>
Optional. This should be the absolute path of a location to which the status feeds
are uploaded. This location should be in the same computer from where data feeds
are fetched. If not specified, the status feeds are uploaded to the same location
as the data feeds. If HTTP is used to fetch the data feed, the value of this tag
should be the HTTP URL to which the status feed can be posted. If this tag is not
specified, the status feed is posted to the HTTP URL of the data feed.
  </xsd:documentation>
 </xsd:annotation>
</xsd:element>

<xsd:element name="securityType" default="noSecurity" maxOccurs="1" minOccurs="0">
 <xsd:simpleType>
  <xsd:restriction base="xsd:string">
   <xsd:enumeration value="identityBased"/>
   <xsd:enumeration value="attributeBased"/>
   <xsd:enumeration value="noSecurity"/>
  </xsd:restriction>
 </xsd:simpleType>
</xsd:element>

<xsd:element name="securityAttribute" minOccurs="0" maxOccurs="unbounded">
 <xsd:complexType>
  <xsd:simpleContent>
   <xsd:extension base="xsd:string">
    <xsd:attribute name="name" type="xsd:string" use="required"/>
    <xsd:attribute name="grant" type="xsd:boolean" default="true"/>
   </xsd:extension>
  </xsd:simpleContent>
 </xsd:complexType>
</xsd:element>

</xsd:sequence>
</xsd:complexType>

</xsd:element>
</xsd:schema>
```

Control Feed Example

The following example shows a control feed used in an XML-connector based source.

```
<?xml version="1.0" encoding="windows-1252" ?>
<rss xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" version="2.0"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema">
 <channel>
   <title>Contacts</title>
   <link>http://my.company.com/rss</link>
   <description>The channel contains feed for contacts</description>
```

```
          <lastBuildDate>2006-04-03T12:20:20.00Z</lastBuildDate>
          <channelDesc xmlns="http://xmlns.oracle.com/orarss">
            <feedType>control</feedType>
          </channelDesc>
          <item>
            <link>file://localhost/C:\project\rss_feeds\test.xml</link>
          </item>
          <item>
            <link>file://localhost/C:\project\rss_feeds\test2.xml</link>
          </item>
          <item>
            <link>http://my.host.com/contacts/control.xml</link>
            <itemDesc xmlns="http://xmlns.oracle.com/orarss"
operation="control"></itemDesc>
          </item>
          <item>
            <link>file://localhost/C:\project\rss_feeds\test3.xml</link>
          </item>
       </channel>
    </rss>
```

Control Feed XML Schema Definition

The following are the XSDs for the control feed.

controlfeed.xsd

```xml
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
 xmlns:oss="http://xmlns.oracle.com/orarss"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
elementFormDefault="qualified">
<xsd:import namespace="http://xmlns.oracle.com/orarss" schemaLocation="orarss_
controlfeed.xsd" />
  <xsd:element name="rss">
    <xsd:annotation>
      <xsd:documentation>RSS control feed</xsd:documentation>
    </xsd:annotation>
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="channel" type="rssChannelType"/>
      </xsd:sequence>
      <xsd:attribute name="version" type="xsd:decimal" use="required"
fixed="2.0"/>
    </xsd:complexType>
  </xsd:element>
  <xsd:complexType name="rssChannelType">
    <xsd:sequence>
      <xsd:element name="title" type="xsd:string"/>
      <xsd:element name="link" type="xsd:anyURI"/>
      <xsd:element name="description" type="xsd:string"/>
      <xsd:element name="lastBuildDate" type="xsd:dateTime">
        <xsd:annotation>
          <xsd:documentation>
              This is the publishing date for this channel
          </xsd:documentation>
        </xsd:annotation>
      </xsd:element>
      <xsd:any namespace="##other" minOccurs="0"/>
```

```
        <xsd:element ref="oss:channelDesc" />
        <xsd:element name="item" type="itemType" maxOccurs="unbounded"/>
      </xsd:sequence>
    </xsd:complexType>
    <xsd:complexType name="itemType">
      <xsd:all>
        <xsd:element name="link" type="xsd:anyURI"/>
        <xsd:element name="title" type="xsd:string" minOccurs="0"/>
        <xsd:element name="description" type="xsd:string" minOccurs="0">
          <xsd:annotation>
            <xsd:documentation>
              The description is ignored as far as Oracle processing is concerned
            </xsd:documentation>
          </xsd:annotation>
        </xsd:element>
        <xsd:element ref="oss:itemDesc" minOccurs="0"/>
      </xsd:all>
    </xsd:complexType>
</xsd:schema>
```

orarss_controlfeed.xsd

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
targetNamespace="http://xmlns.oracle.com/orarss"
 xmlns:oss="http://xmlns.oracle.com/orarss"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
elementFormDefault="qualified">
  <xsd:element name="channelDesc" type="oss:channelDescType" />
     <xsd:element name="itemDesc">
    <xsd:complexType>
          <xsd:attribute name="operation" type="oss:operationType"/>
    </xsd:complexType>
  </xsd:element>
  <xsd:complexType name="descriptionType" abstract="true"/>
  <xsd:complexType name="channelDescType">
    <xsd:complexContent>
      <xsd:extension base="oss:descriptionType">
        <xsd:all>
          <xsd:element name="sourceName" type="xsd:string" minOccurs="0">
            <xsd:annotation>
              <xsd:documentation>
The business Object that this channel corresponds to - if missing then the channel
contains information from multiple objects.
              </xsd:documentation>
            </xsd:annotation>
          </xsd:element>
          <xsd:element name="feedType" maxOccurs="1" minOccurs="0">
            <xsd:simpleType>
              <xsd:restriction base="xsd:string">
                <xsd:enumeration value="control"/>
              </xsd:restriction>
            </xsd:simpleType>
          </xsd:element>
          <xsd:element name="batchId" type="xsd:string" minOccurs="0"/>
          <xsd:element name="itemCount" type="xsd:positiveInteger"
minOccurs="0"/>
        </xsd:all>
      </xsd:extension>
    </xsd:complexContent>
```

```
</xsd:complexType>
<xsd:simpleType name="operationType">
    <xsd:restriction base="xsd:string">
      <xsd:enumeration value="control"/>
    </xsd:restriction>
</xsd:simpleType>
</xsd:schema>
```

Data Feed Example

When specifying a date-time value in an attribute for an RSS feed, use this format:

*YYYY-MM-DD*T*hh:mm:ss*tz[+|-*offset*]

Where:

YYYY is the year (4 digits).

MM is the month (2 digits).

DD is the day (2 digits).

T is a literal character used as a separator.

hh is the hour (2 digits in 24-hour time designation).

mm is the minute (2 digits).

ss is the second (2 digits).

tz is an RFC822 time zone. RFC822 supports UT, GMT, and Z for universal time, and the following North American time zones: EST, EDT, CST, CDT, MST, MDT, PST, and PDT.

offset is the difference between the local time zone and GMT in the format *hh:mm*.

All of these examples are correct:

```
2002-06-03T16:06:05.00GMT
2002-06-03T16:06:05.00UT
2002-06-03T16:06:05.00PDT
2002-06-03T16:06:05.00GMT-07:00
2002-06-03T16:06:05.00JST
2002-06-03T16:06:05.00GMT+09:00
2002-06-03T16:06:05.00Z
```

As shown by these date-time examples, you can specify numeric offsets in the date-time string. Thus, Greenwich Mean Time can be specified as either GMT or GMT+00:00 or UT or Z, and Eastern Standard Time can be specified as either EST or GMT-05:00.

Repositories that dispatch data feeds must explicitly specify the content type to avoid the filtering overhead required by Oracle SES to detect the content type. See the <contentLink> and <content> elements in the example.

The following example shows a data feed containing three documents.

```
<?xml version="1.0" encoding="UTF-8"?>
<rss xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  version="2.0" xmlns:xsd="http://www.w3.org/2001/XMLSchema">
<channel>
<title>Contacts</title>
<link>http://my.company.com/rss</link>
<description>The channel contains feed for contacts</description>
```

```
<lastBuildDate>2006-04-03T12:20:20.00Z</lastBuildDate>
<channelDesc xmlns="http://xmlns.oracle.com/orarss">
    <feedType>full</feedType>
</channelDesc>
<item>
    <link>http://my.company.com/contacts?id=paul</link>
    <itemDesc xmlns="http://xmlns.oracle.com/orarss" operation="insert">
    <documentMetadata>
      <author>Administrator</author>
      <accessURL>http://foo.com</accessURL>
      <lastModifiedDate>2009-12-12T12:22:22.00Z</lastModifiedDate>
      <keywords>Content Contact</keywords>
      <summary>This is the summary of the document.</summary>
      <sourceHierarchy>
          <path>company</path>
          <path>department</path>
          <path>group</path>
      </sourceHierarchy>
      <docAttr name="organization">Reports</docAttr>
      <docAttr name="country">Germany</docAttr>
    </documentMetadata>
    <documentAcl>
      <securityAttr name="EMPLOYEE_ID">0R9NH</securityAttr>
    </documentAcl>
    <documentInfo>
      <status>STATUS_OK_FOR_INDEX</status>
    </documentInfo>
    <documentContent>
      <contentLink
contentType="text/html">http://my.company.com/reports.html</contentLink>
      <content contentType="text/plain">Paul Robinson, A240, Westland
Drive</content>
    </documentContent>
  </itemDesc>
</item>
<item>
   <link>http://my.company.com/contacts?id=tom</link>
   <itemDesc xmlns="http://xmlns.oracle.com/orarss" operation="delete"/>
</item>
<item>
   <link>http://my.company.com/contacts?id=robert</link>
   <itemDesc xmlns="http://xmlns.oracle.com/orarss" operation="insert">
   <documentMetadata>
     <author>Administrator</author>
     <accessURL>http://foo.com</accessURL>
     <lastModifiedDate>2009-12-12T12:22:22.00Z</lastModifiedDate>
     <keywords>Content Contact </keywords>
     <summary>This is the summary of the document</summary>
     <sourceHierarchy>
       <path>company</path>
       <path>department</path>
       <path>group</path>
     </sourceHierarchy>
     <docAttr name="organization">Sales</docAttr>
     <docAttr name="country">China</docAttr>
     </documentMetadata>
     <documentAcl>
       <securityAttr name="EMPLOYEE_ID">I23489</securityAttr>
     </documentAcl>
     <documentInfo>
```

```
      <status>STATUS_OK_FOR_INDEX</status>
    </documentInfo>
    <documentContent>
      <contentLink
contentType="text/html">http://my.company.com/sales.html</contentLink >
      <content contentType="text/plain">Robert Mogambo, C318, Lakeside
Avenue</content>
    </documentContent>
    </itemDesc>
</item>
</channel>
</rss>
```

Data Feed XML Schema Definition

The following are the XSDs for the data feed.

datafeed.xsd

```
<?xml version="1.0" encoding="windows-1252"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
 xmlns:oss="http://xmlns.oracle.com/orarss"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
elementFormDefault="qualified">
<xsd:import namespace="http://xmlns.oracle.com/orarss" schemaLocation="orarss_
datafeed.xsd" />
  <xsd:element name="rss">
    <xsd:annotation>
      <xsd:documentation>RSS data feed</xsd:documentation>
    </xsd:annotation>
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="channel" type="rssChannelType"/>
      </xsd:sequence>
      <xsd:attribute name="version" type="xsd:decimal" use="required"
fixed="2.0"/>
    </xsd:complexType>
  </xsd:element>

  <xsd:complexType name="rssChannelType">
    <xsd:sequence>
      <xsd:element name="title" type="xsd:string"/>
      <xsd:element name="link" type="xsd:anyURI"/>
      <xsd:element name="description" type="xsd:string"/>
      <xsd:element name="lastBuildDate" type="xsd:dateTime">
        <xsd:annotation>
          <xsd:documentation>
            This is the publishing date for this channel
          </xsd:documentation>
        </xsd:annotation>
      </xsd:element>
      <xsd:element ref="oss:channelDesc" />
      <xsd:element name="item" type="itemType"  maxOccurs="unbounded" />
    </xsd:sequence>
  </xsd:complexType>

  <xsd:complexType name="itemType">
    <xsd:all>
      <xsd:element name="link" type="xsd:anyURI">
```

```
          <xsd:annotation>
            <xsd:documentation>
              Display URL of the item. This URL should be UTF-8 encoded.
            </xsd:documentation>
          </xsd:annotation>
        </xsd:element>
        <xsd:element name="title" type="xsd:string" minOccurs="0"/>
        <xsd:element name="description" type="xsd:string" minOccurs="0"/>
        <xsd:element ref="oss:itemDesc"/>
      </xsd:all>
    </xsd:complexType>
</xsd:schema>
```

orarss_datafeed.xsd

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
targetNamespace="http://xmlns.oracle.com/orarss"
 xmlns:oss="http://xmlns.oracle.com/orarss"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
elementFormDefault="qualified">
  <xsd:element name="channelDesc" type="oss:channelDescType"/>
  <xsd:element name="itemDesc">
    <xsd:complexType>
      <xsd:complexContent>
        <xsd:extension base="oss:itemDescType">
          <xsd:attribute name="operation" type="oss:operationType"
default="insert"/>
        </xsd:extension>
      </xsd:complexContent>
    </xsd:complexType>
  </xsd:element>
  <xsd:complexType name="descriptionType" abstract="true"/>
  <xsd:complexType name="channelDescType">
    <xsd:complexContent>
      <xsd:extension base="oss:descriptionType">
        <xsd:all>
          <xsd:element name="sourceName" type="xsd:string" minOccurs="0">
            <xsd:annotation>
              <xsd:documentation>
The business Object that this channel corresponds to - if missing then the channel
contains information from multiple objects.
              </xsd:documentation>
            </xsd:annotation>
          </xsd:element>
          <xsd:element name="feedType" default="incremental" maxOccurs="1"
minOccurs="0">
            <xsd:simpleType>
              <xsd:restriction base="xsd:string">
                <xsd:enumeration value="incremental"/>
                <xsd:enumeration value="full"/>
              </xsd:restriction>
            </xsd:simpleType>
          </xsd:element>
          <xsd:element name="batchId" type="xsd:string" minOccurs="0"/>
          <xsd:element name="itemCount" type="xsd:positiveInteger"
minOccurs="0"/>
        </xsd:all>
      </xsd:extension>
    </xsd:complexContent>
```

```
          </xsd:complexType>
   <xsd:complexType name="itemDescType">
     <xsd:complexContent>
       <xsd:extension base="oss:descriptionType">
         <xsd:all>
           <xsd:element name="documentMetadata" type="oss:metadataType"
minOccurs="0"/>
           <xsd:element name="documentAcl" type="oss:securityType" minOccurs="0"/>
           <xsd:element name="documentInfo" type="oss:infoType" minOccurs="0"/>
           <xsd:element name="documentContent" type="oss:bodyType" minOccurs="0"/>
         </xsd:all>
       </xsd:extension>
     </xsd:complexContent>
   </xsd:complexType>
   <xsd:simpleType name="operationType">
     <xsd:restriction base="xsd:string">
       <xsd:enumeration value="insert"/>
       <xsd:enumeration value="replace"/>
       <xsd:enumeration value="delete"/>
     </xsd:restriction>
   </xsd:simpleType>
   <xsd:group name="authorgroup">
     <xsd:sequence>
     </xsd:sequence>
   </xsd:group>
   <xsd:group name="docattrgroup">
       <xsd:sequence>
       </xsd:sequence>
   </xsd:group>
   <xsd:complexType name="metadataType">
     <xsd:choice minOccurs="0" maxOccurs="unbounded">
       <xsd:element name="author" type="xsd:string" minOccurs="0"
maxOccurs="unbounded" />
       <xsd:element name="accessURL" type="xsd:string" minOccurs="0"/>
       <xsd:element name="lastModifiedDate" type="xsd:string" minOccurs="0"/>
       <xsd:element name="keywords" type="xsd:string" minOccurs="0"/>
       <xsd:element name="summary" type="xsd:string" minOccurs="0"/>
       <xsd:element name="language" type="xsd:string" minOccurs="0"/>
       <xsd:element name="sourceHierarchy" type="oss:sourceHierType"
minOccurs="0"/>
       <xsd:element name="docAttr" minOccurs="0" maxOccurs="unbounded">
         <xsd:complexType>
         <xsd:simpleContent>
         <xsd:extension base="oss:docAttrType">
         <xsd:attribute name="name" type="xsd:string" use="required" />
         <xsd:attribute name="type" use="optional" default="string">
         <xsd:simpleType>
         <xsd:restriction base="xsd:string">
         <xsd:enumeration value="string"></xsd:enumeration>
         <xsd:enumeration value="decimal"></xsd:enumeration>
         <xsd:enumeration value="dateTime"></xsd:enumeration>
         </xsd:restriction>
         </xsd:simpleType>
         </xsd:attribute>
         </xsd:extension>
         </xsd:simpleContent>
         </xsd:complexType>
       </xsd:element>
     </xsd:choice>
   </xsd:complexType>
```

```
<xsd:complexType name="sourceHierType">
  <xsd:sequence>
    <xsd:element name="path" type="xsd:string" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>
<xsd:simpleType name="statusType">
  <xsd:restriction base="xsd:string">
    <xsd:enumeration value="STATUS_OK_FOR_INDEX"/>
    <xsd:enumeration value="STATUS_BAD_REQUEST"/>
    <xsd:enumeration value="STATUS_AUTH_REQUIRED"/>
    <xsd:enumeration value="STATUS_ACCESS_FORBIDDEN"/>
    <xsd:enumeration value="STATUS_NOTFOUND"/>
    <xsd:enumeration value="STATUS_PROXY_REQUIRED"/>
    <xsd:enumeration value="STATUS_REQUEST_TIMEOUT"/>
    <xsd:enumeration value="STATUS_SERVER_ERROR"/>
    <xsd:enumeration value="STATUS_BAD_GATEWAY"/>
    <xsd:enumeration value="STATUS_FETCH_ERROR"/>
    <xsd:enumeration value="STATUS_READ_TIMEOUT"/>
    <xsd:enumeration value="STATUS_FILTER_ERROR"/>
    <xsd:enumeration value="STATUS_OUT_OF_MEMORY"/>
    <xsd:enumeration value="STATUS_IO_EXCEPTION"/>
    <xsd:enumeration value="STATUS_CONNECTION_REFUSED"/>
    <xsd:enumeration value="STATUS_DUPLICATE_DOC"/>
    <xsd:enumeration value="STATUS_EMPTY_DOC"/>
    <xsd:enumeration value="STATUS_LOGIN_FAILED"/>
    <xsd:enumeration value="STATUS_OK_BUT_NO_INDEX"/>
    <xsd:enumeration value="STATUS_OK_CRAWLED"/>
    <xsd:enumeration value="STATUS_CANNOT_READ"/>
    <xsd:enumeration value="STATUS_DOC_SIZE_TOO_BIG"/>
  </xsd:restriction>
</xsd:simpleType>
<xsd:simpleType name="docAttrType">
  <xsd:union memberTypes="xsd:decimal xsd:string"/>
</xsd:simpleType>

<xsd:complexType name="infoType">
  <xsd:sequence>
    <xsd:element name="status" type="oss:statusType"/>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="securityType">
  <xsd:choice>
    <xsd:element name="principal" minOccurs="0" maxOccurs="unbounded">
      <xsd:complexType>
        <xsd:simpleContent>
          <xsd:extension base="xsd:string">
            <xsd:attribute name="idType" type="oss:idAttrType" use="optional"
default="user">
              <xsd:annotation>
                <xsd:documentation>
                  User or group.
                </xsd:documentation>
              </xsd:annotation>
            </xsd:attribute>
            <xsd:attribute name="format" type="xsd:string" use="required"/>
            <xsd:attribute name="grant" type="xsd:boolean" use="optional"
default="true"/>
          </xsd:extension>
        </xsd:simpleContent>
```

```
          </xsd:complexType>
        </xsd:element>
        <xsd:element name="securityAttr" minOccurs="0" maxOccurs="unbounded">
          <xsd:complexType>
            <xsd:simpleContent>
              <xsd:extension base="xsd:string">
                <xsd:attribute name="name" type="xsd:string" use="required"/>
              </xsd:extension>
            </xsd:simpleContent>
          </xsd:complexType>
        </xsd:element>
      </xsd:choice>
      <xsd:attribute name="ownerGuid" type="xsd:string">
        <xsd:annotation>
          <xsd:documentation>
            OwnerGUID - useful if the principal includes OWNER
          </xsd:documentation>
        </xsd:annotation>
      </xsd:attribute>
  </xsd:complexType>

  <xsd:simpleType name="idAttrType">
    <xsd:restriction base="xsd:string">
      <xsd:enumeration value="user"/>
      <xsd:enumeration value="group"/>
      <xsd:enumeration value="owner"/>
    </xsd:restriction>
  </xsd:simpleType>

  <xsd:complexType name="bodyType">
    <xsd:choice minOccurs="0" maxOccurs="unbounded">
      <xsd:element name="contentLink" minOccurs="0" maxOccurs="unbounded">
        <xsd:complexType>
          <xsd:simpleContent>
            <xsd:extension base="xsd:anyURI">
              <xsd:attribute name="contentType" default="text/html"/>
              <xsd:attribute name="inline" type="xsd:string" default="false"/>
            </xsd:extension>
          </xsd:simpleContent>
        </xsd:complexType>
      </xsd:element>
      <xsd:element name="content" minOccurs="0">
        <xsd:complexType>
          <xsd:simpleContent>
            <xsd:extension base="xsd:string">
              <xsd:attribute name="contentType" type="xsd:string"
default="text/plain"/>
              <xsd:attribute name="inline" type="xsd:string" default="true"/>
            </xsd:extension>
          </xsd:simpleContent>
        </xsd:complexType>
      </xsd:element>
      <xsd:element name="xhtmlContent" minOccurs="0" maxOccurs="1">
        <xsd:complexType>
          <xsd:complexContent>
            <xsd:extension base="xsd:anyType">
              <xsd:attribute name="lang" type="xsd:string"/>
              <xsd:anyAttribute namespace="##other"/>
            </xsd:extension>
          </xsd:complexContent>
```

```
            </xsd:complexType>
          </xsd:element>
        </xsd:choice>
      </xsd:complexType>
    </xsd:schema>
```

B

URL Crawler Status Codes

The crawler uses a set of codes to indicate the result of the crawled URL. Besides the standard HTTP status code, it uses its own code for non-HTTP related situations.

Only URLs with status 200 are indexed. If the record exists in `EQ$URL` but the status is something other than 200, then the crawler encountered an error trying to fetch the document. A status of less than 600 maps directly to the HTTP status code.

> **See Also:** "Status Code Definitions" in *Hypertext Transfer Protocol -- HTTP/1.1* at
>
> http://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html

The following table lists the URL status codes, document container codes used by the crawler plug-in, and EQG codes.

Code	Description	Document Container Code	EQG Codes
0	A URL that has been enqueued but not yet processed		N/A
200	URL OK	STATUS_OK_FOR_INDEX	N/A
400	Bad request	STATUS_BAD_REQUEST	30009
401	Authorization required	STATUS_AUTH_REQUIRED	30007
402	Payment required		30011
403	Access forbidden	STATUS_ACCESS_FORBIDDEN	30010
404	Not found	STATUS_NOTFOUND	30008
405	Method not allowed		30012
406	Not acceptable		30013
407	Proxy authentication required	STATUS_PROXY_REQUIRED	30014
408	Request timeout	STATUS_REQUEST_TIMEOUT	30015
409	Conflict		30016
410	Gone		30017
414	Request URI too large		30066
500	Internal server error	STATUS_SERVER_ERROR	10018
501	Not implemented		10019
502	Bad gateway	STATUS_BAD_GATEWAY	10020

Code	Description	Document Container Code	EQG Codes
503	Service unavailable	STATUS_FETCH_ERROR	10021
504	Gateway timeout		10022
505	HTTP version not supported		10023
902	Timeout reading document	STATUS_READ_TIMEOUT	30057
903	Filtering failed	STATUS_FILTER_ERROR	30065
904	Out of memory error	STATUS_OUT_OF_MEMORY	30003
905	IOEXCEPTION in processing URL	STATUS_IO_EXCEPTION	30002
906	Connection refused	STATUS_CONNECTION_REFUSED	30025
907	Socket bind exception		30079
908	Filter not available		30081
909	Duplicate document detected		30082
910	Duplicate document ignored	STATUS_DUPLICATE_DOC	30083
911	Empty document	STATUS_EMPTY_DOC	30106
951	URL not indexed (this can happen if robots.txt specifies that a certain document should not be indexed)	STATUS_OK_BUT_NO_INDEX	N/A
952	URL crawled	STATUS_OK_CRAWLED	N/A
953	Metatag redirection		N/A
954	HTTP redirection		30000
955	Black list URL		N/A
956	URL is not unique		31017
957	Sentry URL (URL as a place holder)		N/A
958	Document read error	STATUS_CANNOT_READ	30173
959	Form login failed	STATUS_LOGIN_FAILED	30183
960	Document size too big, ignored	STATUS_DOC_SIZE_TOO_BIG	30209
962	Document was excluded based on mime type	STATUS_DOC_MIME_TYPE_ EXCLUDED	30041
964	Document was excluded based on boundary rules	STATUS_DOC_BOUNDARY_RULE_ EXCLUDED	30258
1001	Datatype is not TEXT/HTML		30001
1002	Broken network data stream		30004
1003	HTTP redirect location does not exist		30005
1004	Bad relative URL		30006
1005	HTTP error		30024
1006	Error parsing HTTP header		30058
1007	Invalid URL table column name		30067
1009	Binary document reported as text document		30126
1010	Invalid display URL		30112

Code	Description	Document Container Code	EQG Codes
1011	Invalid XML from OracleAS Portal	`PORTAL_XMLURL_FAIL`	31011
1020-1024	URL is not reachable. The status starts at 1020, and it increases by one with each try. After five tries (if it reaches 1025), the URL is deleted.		N/A
1026-1029	URL cannot be found. The status turns from 404 to 1026 when a URL cannot be found on re-crawl, and it increases by one with each try. After five tries (if it reaches 1030), the URL is deleted.		N/A
1111	URL remained in the queue even after a successful crawl. This indicates that the crawler had a problem processing this document. You could investigate the URL by crawling it in a separate source to check for errors in the crawler log.		N/A

C

Third Party Licenses

This appendix includes the third party licenses for all the third party products included with Oracle Secure Enterprise Search. This appendix includes the following topics:

- Apache Software
- Eclipse Software
- Egothor Software
- Javascript Bubbling Library
- Oracle Enterprise Scheduler
- Plug-in Software
- J2SSH Software Version 0.2.7
- Snowball Software
- Visigoth Software
- Yahoo! Inc.

Apache Software

This product contains code from the Apache Software Foundation ("Apache"). Under the terms of the Apache license, Oracle is required to provide the following notices. Note, however, that the Oracle program license that accompanied this product determines your right to use the Oracle program, including the Apache software, and the terms contained in the following notices do not change those rights. Notwithstanding anything to the contrary in the Oracle program license, the Apache software is provided by Oracle "AS IS" and without any warranty or support of any kind from Oracle or Apache.

```
              Apache License
         Version 2.0, January 2004
      http://www.apache.org/licenses/


TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION
1. Definitions.

     "License" shall mean the terms and conditions for use, reproduction,
     and distribution as defined by Sections 1 through 9 of this document.

     "Licensor" shall mean the copyright owner or entity authorized by
     the copyright owner that is granting the License.
```

"Legal Entity" shall mean the union of the acting entity and all
other entities that control, are controlled by, or are under common
control with that entity. For the purposes of this definition,
"control" means (i) the power, direct or indirect, to cause the
direction or management of such entity, whether by contract or
otherwise, or (ii) ownership of fifty percent (50%) or more of the
outstanding shares, or (iii) beneficial ownership of such entity.

"You" (or "Your") shall mean an individual or Legal Entity
exercising permissions granted by this License.

"Source" form shall mean the preferred form for making modifications,
including but not limited to software source code, documentation
source, and configuration files.

"Object" form shall mean any form resulting from mechanical
transformation or translation of a Source form, including but
not limited to compiled object code, generated documentation,
and conversions to other media types.

"Work" shall mean the work of authorship, whether in Source or
Object form, made available under the License, as indicated by a
copyright notice that is included in or attached to the work
(an example is provided in the Appendix below).

"Derivative Works" shall mean any work, whether in Source or Object
form, that is based on (or derived from) the Work and for which the
editorial revisions, annotations, elaborations, or other modifications
represent, as a whole, an original work of authorship. For the purposes
of this License, Derivative Works shall not include works that remain
separable from, or merely link (or bind by name) to the interfaces of,
the Work and Derivative Works thereof.

"Contribution" shall mean any work of authorship, including
the original version of the Work and any modifications or additions
to that Work or Derivative Works thereof, that is intentionally
submitted to Licensor for inclusion in the Work by the copyright owner
or by an individual or Legal Entity authorized to submit on behalf of
the copyright owner. For the purposes of this definition, "submitted"
means any form of electronic, verbal, or written communication sent
to the Licensor or its representatives, including but not limited to
communication on electronic mailing lists, source code control systems,
and issue tracking systems that are managed by, or on behalf of, the
Licensor for the purpose of discussing and improving the Work, but
excluding communication that is conspicuously marked or otherwise
designated in writing by the copyright owner as "Not a Contribution."

"Contributor" shall mean Licensor and any individual or Legal Entity
on behalf of whom a Contribution has been received by Licensor and
subsequently incorporated within the Work.

2. Grant of Copyright License. Subject to the terms and conditions of
   this License, each Contributor hereby grants to You a perpetual,
   worldwide, non-exclusive, no-charge, royalty-free, irrevocable
   copyright license to reproduce, prepare Derivative Works of,
   publicly display, publicly perform, sublicense, and distribute the
   Work and such Derivative Works in Source or Object form.

3. Grant of Patent License. Subject to the terms and conditions of
   this License, each Contributor hereby grants to You a perpetual,

worldwide, non-exclusive, no-charge, royalty-free, irrevocable
(except as stated in this section) patent license to make, have made,
use, offer to sell, sell, import, and otherwise transfer the Work,
where such license applies only to those patent claims licensable
by such Contributor that are necessarily infringed by their
Contribution(s) alone or by combination of their Contribution(s)
with the Work to which such Contribution(s) was submitted. If You
institute patent litigation against any entity (including a
cross-claim or counterclaim in a lawsuit) alleging that the Work
or a Contribution incorporated within the Work constitutes direct
or contributory patent infringement, then any patent licenses
granted to You under this License for that Work shall terminate
as of the date such litigation is filed.

4. Redistribution. You may reproduce and distribute copies of the
Work or Derivative Works thereof in any medium, with or without
modifications, and in Source or Object form, provided that You
meet the following conditions:

(a) You must give any other recipients of the Work or
Derivative Works a copy of this License; and

(b) You must cause any modified files to carry prominent notices
stating that You changed the files; and

(c) You must retain, in the Source form of any Derivative Works
that You distribute, all copyright, patent, trademark, and
attribution notices from the Source form of the Work,
excluding those notices that do not pertain to any part of
the Derivative Works; and

(d) If the Work includes a "NOTICE" text file as part of its
distribution, then any Derivative Works that You distribute must
include a readable copy of the attribution notices contained
within such NOTICE file, excluding those notices that do not
pertain to any part of the Derivative Works, in at least one
of the following places: within a NOTICE text file distributed
as part of the Derivative Works; within the Source form or
documentation, if provided along with the Derivative Works; or,
within a display generated by the Derivative Works, if and
wherever such third-party notices normally appear. The contents
of the NOTICE file are for informational purposes only and
do not modify the License. You may add Your own attribution
notices within Derivative Works that You distribute, alongside
or as an addendum to the NOTICE text from the Work, provided
that such additional attribution notices cannot be construed
as modifying the License.

You may add Your own copyright statement to Your modifications and
may provide additional or different license terms and conditions
for use, reproduction, or distribution of Your modifications, or
for any such Derivative Works as a whole, provided Your use,
reproduction, and distribution of the Work otherwise complies with
the conditions stated in this License.

5. Submission of Contributions. Unless You explicitly state otherwise,
any Contribution intentionally submitted for inclusion in the Work
by You to the Licensor shall be under the terms and conditions of
this License, without any additional terms or conditions.
Notwithstanding the above, nothing herein shall supersede or modify

the terms of any separate license agreement you may have executed
with Licensor regarding such Contributions.

6. Trademarks. This License does not grant permission to use the trade
   names, trademarks, service marks, or product names of the Licensor,
   except as required for reasonable and customary use in describing the
   origin of the Work and reproducing the content of the NOTICE file.

7. Disclaimer of Warranty. Unless required by applicable law or
   agreed to in writing, Licensor provides the Work (and each
   Contributor provides its Contributions) on an "AS IS" BASIS,
   WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or
   implied, including, without limitation, any warranties or conditions
   of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A
   PARTICULAR PURPOSE. You are solely responsible for determining the
   appropriateness of using or redistributing the Work and assume any
   risks associated with Your exercise of permissions under this License.

8. Limitation of Liability. In no event and under no legal theory,
   whether in tort (including negligence), contract, or otherwise,
   unless required by applicable law (such as deliberate and grossly
   negligent acts) or agreed to in writing, shall any Contributor be
   liable to You for damages, including any direct, indirect, special,
   incidental, or consequential damages of any character arising as a
   result of this License or out of the use or inability to use the
   Work (including but not limited to damages for loss of goodwill,
   work stoppage, computer failure or malfunction, or any and all
   other commercial damages or losses), even if such Contributor
   has been advised of the possibility of such damages.

9. Accepting Warranty or Additional Liability. While redistributing
   the Work or Derivative Works thereof, You may choose to offer,
   and charge a fee for, acceptance of support, warranty, indemnity,
   or other liability obligations and/or rights consistent with this
   License. However, in accepting such obligations, You may act only
   on Your own behalf and on Your sole responsibility, not on behalf
   of any other Contributor, and only if You agree to indemnify,
   defend, and hold each Contributor harmless for any liability
   incurred by, or claims asserted against, such Contributor by reason
   of your accepting any such warranty or additional liability.

END OF TERMS AND CONDITION

Eclipse Software

Eclipse Public License v 1.0, 1.0.2, 1.2

THE ACCOMPANYING PROGRAM IS PROVIDED UNDER THE TERMS OF THIS
ECLIPSE PUBLIC LICENSE ("AGREEMENT"). ANY USE, REPRODUCTION OR
DISTRIBUTION OF THE PROGRAM CONSTITUTES RECIPIENT'S ACCEPTANCE
OF THIS AGREEMENT.

1. DEFINITIONS

"Contribution" means:

   a) in the case of the initial Contributor, the initial code and documentation
distributed under this Agreement, and

   b) in the case of each subsequent Contributor:

i) changes to the Program, and

ii) additions to the Program;

where such changes and/or additions to the Program originate from and are distributed by that particular Contributor. A Contribution 'originates' from a Contributor if it was added to the Program by such Contributor itself or anyone acting on such Contributor's behalf. Contributions do not include additions to the Program which: (i) are separate modules of software distributed in conjunction with the Program under their own license agreement, and (ii) are not derivative works of the Program.

"Contributor" means any person or entity that distributes the Program.

"Licensed Patents " mean patent claims licensable by a Contributor which are necessarily infringed by the use or sale of its Contribution alone or when combined with the Program.

"Program" means the Contributions distributed in accordance with this Agreement.

"Recipient" means anyone who receives the Program under this Agreement, including all Contributors.

2. GRANT OF RIGHTS

a) Subject to the terms of this Agreement, each Contributor hereby grants Recipient a non-exclusive, worldwide, royalty-free copyright license to reproduce, prepare derivative works of, publicly display, publicly perform, distribute and sublicense the Contribution of such Contributor, if any, and such derivative works, in source code and object code form.

b) Subject to the terms of this Agreement, each Contributor hereby grants Recipient a non-exclusive, worldwide, royalty-free patent license under Licensed Patents to make, use, sell, offer to sell, import and otherwise transfer the Contribution of such Contributor, if any, in source code and object code form. This patent license shall apply to the combination of the Contribution and the Program if, at the time the Contribution is added by the Contributor, such addition of the Contribution causes such combination to be covered by the Licensed Patents. The patent license shall not apply to any other combinations which include the Contribution. No hardware per se is licensed hereunder.

c) Recipient understands that although each Contributor grants the licenses to its Contributions set forth herein, no assurances are provided by any Contributor that the Program does not infringe the patent or other intellectual property rights of any other entity. Each Contributor disclaims any liability to Recipient for claims brought by any other entity based on infringement of intellectual property rights or otherwise. As a condition to exercising the rights and licenses granted hereunder, each Recipient hereby assumes sole responsibility to secure any other intellectual property rights needed, if any. For example, if a third party patent license is required to allow Recipient to distribute the Program, it is Recipient's responsibility to acquire that license before distributing the Program.

d) Each Contributor represents that to its knowledge it has sufficient copyright rights in its Contribution, if any, to grant the copyright license set forth in this Agreement.

3. REQUIREMENTS

A Contributor may choose to distribute the Program in object code form under its own license agreement, provided that:

a) it complies with the terms and conditions of this Agreement; and

b) its license agreement:

i) effectively disclaims on behalf of all Contributors all warranties and conditions, express and implied, including warranties or conditions of title and non-infringement, and implied warranties or conditions of merchantability and fitness for a particular purpose;

ii) effectively excludes on behalf of all Contributors all liability for damages, including direct, indirect, special, incidental and consequential damages, such as lost profits;

iii) states that any provisions which differ from this Agreement are offered by that Contributor alone and not by any other party; and

iv) states that source code for the Program is available from such Contributor, and informs licensees how to obtain it in a reasonable manner on or through a medium customarily used for software exchange.

When the Program is made available in source code form:

a) it must be made available under this Agreement; and

b) a copy of this Agreement must be included with each copy of the Program.

Contributors may not remove or alter any copyright notices contained within the Program.

Each Contributor must identify itself as the originator of its Contribution, if any, in a manner that reasonably allows subsequent Recipients to identify the originator of the Contribution.

4. COMMERCIAL DISTRIBUTION

Commercial distributors of software may accept certain responsibilities with respect to end users, business partners and the like. While this license is intended to facilitate the commercial use of the Program, the Contributor who includes the Program in a commercial product offering should do so in a manner which does not create potential liability for other Contributors. Therefore, if a Contributor includes the Program in a commercial product offering, such Contributor ("Commercial Contributor") hereby agrees to defend and indemnify every other Contributor ("Indemnified Contributor") against any losses, damages and costs (collectively "Losses") arising from claims, lawsuits and other legal actions brought by a third party against the Indemnified Contributor to the extent caused by the acts or omissions of such Commercial Contributor in connection with its distribution of the Program in a commercial product offering. The obligations in this section do not apply to any claims or Losses relating to any actual or alleged intellectual property infringement. In order to qualify, an Indemnified Contributor must: a) promptly notify the Commercial Contributor in writing of such claim, and b) allow the Commercial Contributor to control, and cooperate with the Commercial Contributor in, the defense and any related settlement negotiations. The Indemnified Contributor may participate in any such claim at its own expense.

For example, a Contributor might include the Program in a commercial product offering, Product X. That Contributor is then a Commercial Contributor. If that Commercial Contributor then makes performance claims, or offers warranties related to Product X, those performance claims and warranties are such Commercial Contributor's responsibility alone. Under this section, the Commercial Contributor would have to defend claims against the other Contributors related to those performance claims and warranties, and if a court requires any other Contributor to pay any damages as a result, the Commercial Contributor must pay those damages.

5. NO WARRANTY

EXCEPT AS EXPRESSLY SET FORTH IN THIS AGREEMENT, THE PROGRAM IS PROVIDED ON AN "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, EITHER EXPRESS OR IMPLIED INCLUDING, WITHOUT LIMITATION, ANY WARRANTIES OR CONDITIONS OF TITLE, NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Each Recipient is solely responsible for determining the appropriateness of using and distributing the Program and assumes all risks associated with its exercise of rights under this Agreement, including but not limited to the risks and costs of program errors, compliance with applicable laws, damage to or loss of data, programs or equipment, and unavailability or interruption of operations.

6. DISCLAIMER OF LIABILITY

EXCEPT AS EXPRESSLY SET FORTH IN THIS AGREEMENT, NEITHER RECIPIENT NOR ANY CONTRIBUTORS SHALL HAVE ANY LIABILITY FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING WITHOUT LIMITATION LOST PROFITS), HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OR DISTRIBUTION OF THE PROGRAM OR THE EXERCISE OF ANY RIGHTS GRANTED HEREUNDER, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

7. GENERAL

If any provision of this Agreement is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Agreement, and without further action by the parties hereto, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

If Recipient institutes patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Program itself (excluding combinations of the Program with other software or hardware) infringes such Recipient's patent(s), then such Recipient's rights granted under Section 2(b) shall terminate as of the date such litigation is filed.

All Recipient's rights under this Agreement shall terminate if it fails to comply with any of the material terms or conditions of this Agreement and does not cure such failure in a reasonable period of time after becoming aware of such noncompliance. If all Recipient's rights under this Agreement terminate, Recipient agrees to cease use and distribution of the Program as soon as reasonably practicable. However, Recipient's obligations under this Agreement and any licenses granted by Recipient relating to the Program shall continue and survive.

Everyone is permitted to copy and distribute copies of this Agreement, but in order to avoid inconsistency the Agreement is copyrighted and may only be modified in the following manner. The Agreement Steward reserves the right to publish new versions (including revisions) of this Agreement from time to time. No one other than the Agreement Steward has the right to modify this Agreement. The Eclipse Foundation is the initial Agreement Steward. The Eclipse Foundation may assign the responsibility to serve as the Agreement Steward to a suitable separate entity. Each new version of the Agreement will be given a distinguishing version number. The Program (including Contributions) may always be distributed subject to the version of the Agreement under which it was received. In addition, after a new version of the Agreement is published, Contributor may elect to distribute the Program (including its Contributions) under the new version. Except as expressly stated in Sections 2(a) and 2(b) above, Recipient receives no rights or licenses to the intellectual property of any Contributor under this Agreement, whether expressly, by implication, estoppel or

otherwise. All rights in the Program not expressly granted under this Agreement are reserved.

This Agreement is governed by the laws of the State of New York and the intellectual property laws of the United States of America. No party to this Agreement will bring a legal action under this Agreement more than one year after the cause of action arose. Each party waives its rights to a jury trial in any resulting litigation.

Egothor Software

This product contains software developed by the Egothor Project. Under the terms of the Egothor license, Oracle is required to provide the following notices. Note, however, that the Oracle program license that accompanied this product determines your right to use the Oracle program, including the Egothor software, and the terms contained in the following notices do not change those rights.

http://egothor.sf.net/

Egothor Software License version 1.00. (C) 1997-2004 Leo Galambos,

(C) 2002-2004 "Egothor developers" on behalf of the Egothor Project

This software is copyrighted by the "Egothor developers". If this license applies to a single file or document, the "Egothor developers" are the people or entities mentioned as copyright holders in that file or document. If this license applies to the Egothor project as a whole, the copyright holders are the people or entities mentioned in the file CREDITS. This file can be found in the same location as this license in the distribution.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, the list of contributors, this list of conditions, and the following disclaimer.

Redistributions in binary form must reproduce the above copyright notice, the list of contributors, this list of conditions, and the disclaimer that follows these conditions in the documentation and/or other materials provided with the distribution.

The name "Egothor" must not be used to endorse or promote products derived from this software without prior written permission. For written permission, please contact <Leo.G@seznam.cz <mailto:Leo.G@seznam.cz>>.

Products derived from this software may not be called "Egothor", nor may "Egothor" appear in their name, without prior written permission from <Leo.G@seznam.cz <mailto:Leo.G@seznam.cz>>.

In addition, we request that you include in the end-user documentation provided with the redistribution and/or in the software itself an acknowledgement equivalent to the following:

"This product includes software developed by the Egothor Project.

http://egothor.sf.net/"

THIS SOFTWARE IS PROVIDED ``AS IS'' AND ANY EXPRESSED OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE EGOTHOR PROJECT OR ITS CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS

OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

This software consists of voluntary contributions made by many individuals on behalf of the Egothor Project and was originally created by Leo Galambos <Leo.G@seznam.cz <mailto:Leo.G@seznam.cz>>.

Javascript Bubbling Library

Javascript Bubbling Library
http://caridy.name/legacy/www.bubbling-library.com/

Copyright (c) 2007, Caridy Patiño. All rights reserved.

Redistribution and use of this software in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

To get started using Bubbling Library, simply include the two source files into the head of your document:

```
<!-- YUI Core -->
<script src="/PATH/TO/utilities.js" type="text/javascript"></script>
<!-- Bubbling Library Core -->
<script src="/PATH/TO/bubbling.js" type="text/javascript"></script>
```

Oracle Enterprise Scheduler

Oracle Secure Enterprise Search (Oracle SES) customers are granted the restricted free license for Oracle Enterprise Scheduler. This license includes the ability to schedule and manage crawler schedules as part of the Oracle SES Administration. This includes deploying Oracle Enterprise Scheduler on a separate server where Oracle SES is installed.

Plug-in Software

Oracle SES ships several *plug-ins* to enterprise sources. (Plug-ins allow Oracle SES to crawl and index content in proprietary systems). For some plug-ins to work, additional software may need to be installed and licensed from the respective vendor; for example, EMC Documentum requires Documentum Foundation Classes (DFC), a Java library, to be installed on the computer running Oracle SES.

The following enterprise sources require additional software to be installed on the computer running Oracle SES:

- EMC Documentum Content Server

- Open Text

- Microsoft Exchange

- Microsoft NTFS may require Microsoft .NET 2.0

> **Note:** The Microsoft Exchange connector uses the Jakarta Slide libraries for WebDAV.

See Also:

- Chapter 7, "Configuring Access to Content Management Sources"

- Chapter 8, "Configuring Access to Collaboration Sources"

J2SSH Software Version 0.2.7

J2SSH is dual licensed; you may distribute it under an Apache style license or the LGPL. You may select whichever license suits your needs.

```
/*
 * =============================================================================
 * J2SSH Software License
 * =============================================================================
 *
 *  Copyright (C) 2002-2003 Lee David Painter and Contributors.
 *
 *  Contributions made by:
 *
 *  Brett Smith
 *  Richard Pernavas
 *  Erwin Bolwidt
 *
 * Redistribution and use in source and binary forms, with or without modifica-
 * tion, are permitted provided that the following conditions are met:
 *
 * 1. Redistributions of source code must retain the above copyright notice,
 *    this list of conditions and the following disclaimer.
 *
 * 2. Redistributions in binary form must reproduce the above copyright notice,
 *    this list of conditions and the following disclaimer in the documentation
 *    and/or other materials provided with the distribution.
 *
 * 3. The end-user documentation included with the redistribution, if any, must
 *    include  the following  acknowledgment:  "This product includes  software
 *    developed  by  SSHTools (http://www.sshtools.com/)."
 *    Alternately, this  acknowledgment may  appear in the software itself,  if
```

```
*      and wherever such third-party acknowledgments normally appear.
*
* 4. The names "J2SSH", "Lee David Painter" and "SSHTools"  must not be used to
*    endorse  or promote  products derived  from this  software without  prior
*    written permission. For written permission, please contact
*    lee@sshtools.com.
*
* 5. Products  derived from this software may not  be called "J2SSH" or
*    "SSHTools", nor may these names appear in their name,  without prior
*    written permission  of Lee David Painter.
*
* THIS SOFTWARE IS PROVIDED ``AS IS'' AND ANY EXPRESSED OR IMPLIED WARRANTIES,
* INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND
* FITNESS  FOR A PARTICULAR  PURPOSE ARE  DISCLAIMED. IN NO  EVENT SHALL  THE
* LEE DAVID PAINTER, SSHTOOLS OR ITS CONTRIBUTORS  BE LIABLE FOR  ANY DIRECT,
* INDIRECT, INCIDENTAL, SPECIAL,  EXEMPLARY, OR CONSEQUENTIAL  DAMAGES (INCLU-
* DING, BUT NOT LIMITED TO, PROCUREMENT  OF SUBSTITUTE GOODS OR SERVICES; LOSS
* OF USE, DATA, OR  PROFITS; OR BUSINESS  INTERRUPTION)  HOWEVER CAUSED AND ON
* ANY  THEORY OF LIABILITY,  WHETHER  IN CONTRACT,  STRICT LIABILITY,  OR TORT
* (INCLUDING  NEGLIGENCE OR  OTHERWISE) ARISING IN  ANY WAY OUT OF THE  USE OF
* THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
*
*
*/
```

Snowball Software

This product contains software developed by Snowball. Under the terms of the Snowball license, Oracle is required to provide the following notices. Note, however, that the Oracle program license that accompanied this product determines your right to use the Oracle program, including the Snowball software, and the terms contained in the following notices do not change those rights.

Copyright (c) 2001, Dr. Martin Porter, and (for the Java developments) Copyright (c) 2002, Richard Boulton.

All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution. Neither the name Snowball nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT

LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Visigoth Software

This product contains software developed by the Visigoth Software Society (http://sourceforge.net/projects/visigoths/).

Copyright (c) 2003 The Visigoth Software Society. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. The end-user documentation included with the redistribution, if any, must include the following reacknowledgment: "This product includes software developed by the Visigoth Software Society (http://www.visigoths.org/)." Alternately, this reacknowledgment may appear in the software itself, if and wherever such third-party acknowledgments normally appear.

3. Neither the name "FreeMarker", "Visigoth", nor any of the names of the project contributors may be used to endorse or promote products derived from this software without prior written permission. For written permission, please contact visigoths@visigoths.org.

4. Products derived from this software may not be called "FreeMarker" or "Visigoth" nor may "FreeMarker" or "Visigoth" appear in their names without prior written permission of the Visigoth Software Society.

THIS SOFTWARE IS PROVIDED "AS IS" AND ANY EXPRESSED OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE VISIGOTH SOFTWARE SOCIETY OR ITS CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE,EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

This software consists of voluntary contributions made by many individuals on behalf of the Visigoth Software Society. For more information on the Visigoth Software Society, please see http://sourceforge.net/projects/visigoths/

Yahoo! Inc.

This product contains software developed by Yahoo! Inc. Under the terms of the Yahoo! Inc. license, Oracle is required to provide the following notices. Note, however, that the Oracle program license that accompanied this product determines your right to use the Oracle program, including the Yahoo! Inc. software, and the terms contained in the following notices do not change those rights.

Software License Agreement (BSD License)

Copyright (c) 2006, Yahoo! Inc.

All rights reserved.

Redistribution and use of this software in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

- Neither the name of Yahoo! Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission of Yahoo! Inc.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE. TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

D

Error Messages

The crawler uses a set of messages to log the crawling activities.

The following table lists the most common crawler error messages.

Message ID	Message	Comment	Action
30025	{0}: Connection refused	The Web site refuses the URL access request.	Check the network setup environment of the computer running the crawler.
30027	Not allowed URL: {0}	A URL link violates boundary rules and is discarded.	Confirm that the URL indeed can be ignored.
30030	Malformed URL: {0}	The URL is not properly formed.	Verify the URL.
30031	Excluded by ROBOTS.TXT: {0}	The robots.txt rule from the Web site of the URL does not allow the URL to be crawled.	Configure the crawler to ignore robots rule only when you are managing the target Web site. Use the Home - Sources - Crawling Parameters page.
30040	Ignore URL: {0}	Redirection to this URL is not allowed by boundary rule.	Confirm that the URL indeed should be ignored.
30041	{0}: excluded by MIME type inclusion rule, URL is {1}	The content type of the URL is not in MIME type inclusion list.	Check if the specified content type should be included.
30054	Excessively long URL: {0}	The URL string is too long, and the URL is ignored.	N/A
30057	{0}: timeout reading document	The target Web site is too slow sending page content.	Increase the crawler timeout threshold from the crawler configuration page. The default is 30 seconds.
30083	{0}: Duplicate document ignored	An identical document has been seen before in the same crawl session. This could be an indication of URL looping; that is, a generation of different URLs pointing back to the same page.	Check if the URL is generated correctly. If necessary, disable indexing dynamic URLs. Use the Home - Sources - Crawling Parameters page.

Message ID	Message	Comment	Action
30126	Binary document reported as text document: "{0}"	A binary file has been sent by the Web site as a text document. In most cases, the URL in question is not a binary format text document, like pdf.	Correct the Web site content type setting for the URL, if possible.
30188	Login form not specified for "{0}"	Unable to perform HTML form login, because the name of the form is not set. In general, the name of the form should be automatically set by the crawler.	Identify the URL of the login page, and check whether this is a regular HTML form login page or an OracleAS Single Sign-on login page. Report the problem to Oracle Support.
30199	Encountered an error while responding to the following HTTP authentication request: [{0}]	Unable to authenticate through the target URL.	Verify if the authentication request is basic authentication or digest authentication. Also confirm the provided authentication credentials.
30201	Missing authentication credentials	Authentication data is not available to access the URL.	Check the type of authentication needed and provide it through the source customization page
30206	Ignoring "{0}" due to host (or redirected host) connection problem	The crawler cannot contact the server of the URL.	Verify that the Web site in question is up and try to re-crawl.
30209	Document size ({0}) too big, ignored: {1}	Document size exceeds the default limit of 10 megabytes.	Increase the document size limit on the Global Settings - Crawler Configuration page.
30215	Excluded by crawling depth limit({0}): {1}	Previously crawled URL is excluded due to newly reduced crawling depth limit.	Confirm that the depth limit is correct.
30782	Invalid document attribute {0} - ignored	Some attribute picked up from the document is not defined for the source. It is ignored.	Most likely this is safe to ignore, unless you know that this particular attribute should be defined for this source. In that case, contact Oracle Support.

E

Crawler Performance Metrics

The crawler reports crawler performance metrics information in the crawler log file in case of any crawl performance problem. These metrics messages are logged hourly in the crawler log file and are prefixed with "#%#%#", so as to make it easier to identify them, and to enable extracting them from the log file.

The following table lists the most common crawler performance metrics reported in the crawler log file.

The "ms" suffix in some of the sample metrics values in the table denotes unit of time in milliseconds. The sample metrics values having format "00:00:00" denotes the duration in hours:minutes:seconds.

Crawler Metrics	Sample Value	Description
Average document processing time	143.4 ms	Average time taken for saving document's content into Oracle SES eq$cache table and to update the corresponding URLs. It is a measure of the database performance for saving a cache file.
Average wait time on cache queue	89.01 ms	Average time taken for the caching thread requesting a document from in-memory cache queue till the time when the document is available.
Average caching time	144.7 ms	Average time taken for getting the next document from cache and saving it into the database. It is almost equal to - [document processing time] + [wait time on cache queue] + [other overheads].
Cumulative number of page submit call	100331	Total number of document submissions from connector to the crawler. Each call submits one document for indexing.
Cumulative time for executing page submit	9:49:21	Total time taken by the crawler to process the submitted documents. This includes the time for - filtering, submitting the document cache for caching to the database, and so on. If this time is too long, then it indicates possible bottleneck on the caching thread side, where saving the cache to the database is not fast enough.
Average time (ms) per page submit call	352 ms	Average time taken by the crawler to finish processing one document.

Crawler Metrics	Sample Value	Description
Cum. number of processUrlEntry call	100331	This is a sub-task of "page submit call". This metrics shows the total number of calls to - convert the submitted document into HTML format, perform filtering if necessary, and process attachments, if any.
Cum. processUrlEntry time	5:06:38	Total time taken for processing URLs. It is about 50% of the page submit call.
Average time (ms) per processUrlEntry call	183 ms	Average time taken for processing URLs.
Cum. number of processUrlInfo call	100331	This is a sub-task of "page submit call". This metrics shows the total number of calls for parsing HTML-formatted documents and applying document service processing on them, such as, extracting topics. For non-connector data sources, it includes calls related to extracting attribute values, such as, title from a document.
Cum. processUrlInfo time	4:39:00	Total time taken for executing the crawler tasks - processUrlEntry and processUrlInfo - for making a document ready for indexing. Generally, this time is more than 95% of the page submit time.
Average time (ms) per filtering call	253 ms	Filtering is a sub-task of "processUrlEntry" task, and it denotes the average time taken for filtering a document.
Cum. number of storing attr. values call	100331	Total number of calls made for storing attribute values in database.
Cum. time for storing attr. values	0:06:04	Total time taken for storing attribute values in database.
Average time (ms) per storing attr. values call	3 ms	Average time taken for storing attribute values in database.
Cum. number of set attr. values call	100331	Total number of calls made for setting attribute values.
Cum. time for set attr. values	0:05:25	Total time taken for setting attribute values.
Average time (ms) per set attr. values call	3 ms	Average time taken for setting attribute values.
Cum. number of adding attr. values call	0	Total number of calls made for adding attribute values.
Cum. time for adding attr. values	0 ms	Total time taken for adding attribute values.
Average time (ms) per adding attr. values call	0 ms	Average time taken for adding attribute values.
Cum. number of set 4k text call	100331	Total number of calls made for storing 4KB of text in database.
Cum. time for set 4k text	0:01:13	Total time taken for storing 4KB of text in database.

Crawler Metrics	Sample Value	Description
Average time (ms) per set 4k text call	0 ms	Average time taken for storing a single 4KB of text in database.
Cum. number of storing security attr. values call	100331	Total number of calls made for storing security attribute values in database.
Cum. time for storing security attr. values	0:02:58	Total time taken for storing security attribute values in database.
Average time (ms) per storing security attr. values call	1 ms	Average time taken for storing a security attribute values in database.
Cum. number of insert URL entry call	100331	Total number of calls made for storing URLs in database.
Cum. time for insert URL entry	0:21:51	Total time taken for storing URLs in database.
Average time (ms) per insert URL entry call	13 ms	Average time taken for storing a URL in database.
Cum. number of pagesProcessed call	100331	Total number of pages processed.
Cum. time for executing pagesProcessed	0:01:54	Total time taken for processing pages.
Cum. time for executing page update	0:01:12	Total time taken for updating pages.
Cum. time for executing copy URL stage	0:00:36	Total time taken for copying URLs.
Average time (ms) per pagesProcessed call	1 ms	Average time taken for processing a page.
Average time (ms) per page update call	0 ms	Average time taken for updating a page.
Average time (ms) per URL stage copy call	0 ms	Average time taken for copying a URL.

F

Crawler Troubleshooting

This appendix lists common crawler issues and their solutions.

Crawler Issue	Symptom	Solution

Crawler process hangs when started on a Windows system	On Windows system, crawler process starts without any further progress after displaying the message "Crawling started at mm/dd/yy hh:mm AM|PM" in the crawler log file. For example:
	
	`========= Crawling status =========`
	`[2014-02-27T10:07:04.593-05:00] [search]`
	`[NOTIFICATION]`
	`[EQG-30304] [oracle.search] [tid: Main Thread]`
	`[ecid:0000KHnsbE5FW705Vzt1ic1J3pIO000001, 0]`
	`Crawling started at 2/27/14 10:07 AM.`
	
	The following error message is also displayed in the crawler log file:
	
	`java.io.IOException: Access is denied`
		On Windows system, disable the User Access Control (UAC) setting and restart the crawler process.

Crawler process does not start on a Solaris system.	On Solaris system, crawler process does not start when maximum heap memory size is set to more than 2 GB. This issue occurs because by default Solaris JVM is launched in 32-bit mode, which cannot support more than 2 GB of heap space.	On Solaris system, launch the crawler using 64-bit JVM by using the "-d64" flag for the `CRAWLER_EXEC_PARAMS` parameter in the `mw_home/Oracle_SES1/bin/clexecutor.sh` file. Example: `CRAWLER_EXEC_PARAMS="-d64 -ms70m -mx4g ..."`

Glossary

ACE

An entry in an access control list (ACL). An ACE either grants or denies access to a data source by a particular user.

ACL

A list of access control entries that determines which users have access to a particular data source.

crawl

The process of reading sources and creating the search engine index.

crawler

An Oracle Secure Enterprise Search program that reads sources to create the search engine index.

DN

Distinguished Name. The unique name of a directory entry in Oracle Internet Directory. It includes all the individual names of the parent entries back to the root. The DN tells you exactly where the entry resides in the directory's hierarchy.

DICOM

Digital Imaging and Communications in Medicine (DICOM) is the predominant medical imaging standard for exchanging digital information between medical imaging equipment (such as radiological imaging) and other systems, ensuring inter operability. DICOM images contain rich metadata about the patient, medical equipment, and other medical information. DICOM is a major feature of Oracle Multimedia 11g release.

document

Unit of indexing, returned as one entry in the hitlist. For example, a document could be all the collected information about a person from a human resources system.

duplicate documents

Documents that are identical to each other; that is, they are the exact same size, same content, same title, and so on.

federated search

Oracle SES provides the capability of searching multiple Oracle SES instances with their own document repositories and indexes. It provides a unified framework to search the different document repositories that are crawled, indexed, and maintained

separately. A *federation broker* calls the *federation endpoint* to collect content matching the search criteria for the sources managed at that endpoint.

hitlist

A list of results for a search.

infosource

Data sources that can be browsed based on the source or path of the documents. An infosource hierarchy has this structure: *Source Group > host* or *folder > folder*. An infosource hierarchy provides a count of documents at levels below the top (that is, the Source Group level). A user's access to the infosource structure relies on the user's access to documents stored in the structure.

Users can search documents under a particular path or level by selecting the corresponding node in the infosource browse hierarchy.

index

An Oracle SES structure that is updated after a crawl. It is used to improve performance of searches.

JDBC

The programming API that enables Java applications to access a database through the SQL language. JDBC drivers are written in Java for platform independence but are specific to each database.

LDAP

Lightweight Directory Access Protocol. A standard for representing and accessing user and group profile information.

lexer

A program that breaks the source text into tokens, usually words, in accordance with a specified language.

LOV

List of values.

near duplicate documents

Documents that are similar to each other. They may or may not be identical to each other.

Oracle Application Server (Oracle AS)

Oracle's integrated application server:

- Is standards compliant (J2EE, Web Services, and XML)

- Delivers a comprehensive set of capabilities, including portal, caching, wireless, integration, and personalization

- Provides a single, unified platform for Java and J2EE, Web Services, XML, SQL, and PL/SQL

OracleAS Portal

A component of Oracle Application Server used for the development, deployment, administration, and configuration of enterprise class portals. OracleAS Portal incorporates a portal building framework with self-service publishing features to enable you to create and manage information accessed within your portal.

OracleAS Single Sign-On

A component of Oracle Application Server that enables users to log in to all features of the Oracle AS product suite and to other Web applications, using a single user name and password.

OracleAS Web Cache

A component of Oracle Application Server that improves the performance, scalability, and availability of frequently used Web sites. By storing frequently accessed URLs in memory, Oracle Application Server Web Cache eliminates the need to repeatedly process requests for those URLs on the Web server.

Oracle Content Database

A consolidated, database-centric content management application that provides a comprehensive, integrated solution for file and document life cycle management. Oracle Content Database also offers a comprehensive set of Web services that developers can use to build and enhance content management applications. This book uses the product name Oracle Content Database to mean *both* Oracle Content Database *and* Oracle Content Services.

Oracle Content Server

Formerly known as Stellent Content Server. Oracle Content Server enables users throughout the organization to contribute content from native desktop applications, manage content through rich library services, publish content to web sites or business applications, and access the content with a browser.

Oracle Content Services

See **Oracle Content Database**.

Oracle HTTP Server

The Web server component of Oracle Application Server, built on Apache Web server technology and used to service HTTP requests. Also referred to as OHS in the guide.

Oracle Internet Directory

A repository for storing user credentials and group memberships. By default, the OracleAS Single Sign-On authenticates user credentials against Oracle Internet Directory information about dispersed users and network resources.

Oracle Secure Enterprise Search application

Application for searching the Oracle Secure Enterprise Search index.

Oracle WebLogic Server

Oracle WebLogic Server is a fast and reliable server that is used to build and run enterprise applications and services. It is the middle tier server on which Oracle SES operates.

relevance

The level of match of the search results to the search string.

schedule

The frequency with which each source is crawled.

search

The process of querying the search engine.

searchctl

A utility for starting and stopping the search engine.

search metadata

Information about the sources, crawls, and schedules.

secure search

A type of search that only returns results that the user is allowed to view based on access privileges.

seed URL

The starting point for a crawl.

SOAP

Simple Object Access Protocol. A lightweight, XML-based protocol for exchanging information in a decentralized, distributed environment. SOAP supports different styles of information exchange, including: Remote Procedure Call style (RPC) and Message-oriented exchange.

sources

A source of data to be searched, such as Web sites, database tables, content management repositories, collaboration repositories, or applications.

WebDAV

Web Distributed Authoring and Versioning (WebDAV) is a standard protocol used to provide users with a file system interface to Oracle XML repository over the Internet. The most popular way of accessing a WebDAV server folder is through WebFolders on Microsoft Windows.

WebDAV is an extension to the HTTP 1.1 protocol. It allows clients to perform remote web content authoring through a coherent set of methods, headers, request body formats, and response body formats. WebDAV provides operations to store and retrieve resources, to create and list contents of resource collections, to lock resources for concurrent access in a coordinated manner, and to set and to retrieve resource properties.

WSDL

A general purpose XML language for describing the interface, protocol bindings, and deployment details of Web services.

Index

crawler log file,   12-12, 13-47

M

mailing list sources
  tips,   6-10
master encryption key,   11-52
metadata,   4-24
Microsoft SharePoint sources,   7-11

O

OAM, *See* Oracle Access Manager
OC4J server,   13-3, 13-5
open_cursors parameter,   12-22
optimizing
  index,   12-17, 12-18
Oracle Access Manager,   11-29
Oracle Calendar sources
  secure,   8-19
Oracle Content Database sources,   7-18
  tips,   7-18
Oracle Content Services,   7-18
Oracle HTTP Server
  channel with Oracle SES,   11-13
  front-ending,   11-13, 11-22
  SSL-protect,   6-16
  with AJP13 port,   6-16
Oracle Internet Directory
  identity plug-in,   11-12
    restrictions,   11-11
  IDM systems,   6-15
  login attribute,   8-20
  overview,   11-13
Oracle RAC
  tuning,   12-20
Oracle Secure Enterprise Search
  accessing Application Server Control
    Console,   12-26
  Administration GUI,   1-3
  components,   1-3
  crawler,   1-4, 4-1
  error messages,   D-1
  getting started,   2-1
  global settings,   2-4
  integration with Oracle Internet Directory,   11-13
  overview,   1-1
  security,   11-1
  statistics,   2-3
  third party licenses
    Apache Axis,   C-1
    Apache log4j,   C-1
  tuning crawl performance,   12-5
  what's new in 10.1.7,   xxi
Oracle undo space,   12-23
ORACLE_BASE,   xviii, xix
ORACLE_HOME,   xviii
OracleAS Portal sources,   11-5
  tips,   6-13
  user privileges,   6-13

OracleAS Single Sign-On,   11-5, 11-13

P

parallel querying,   12-15
partitioning,   12-15
passwords
  temporary,   11-5
path rules,   4-4
preference profiles,   3-16
processes parameter,   12-22
push crawling,   4-30
Push Feed,   4-30

Q

Query,   3-1
query application,   3-1, 3-3, 5-4, 5-7, 5-22, 11-13
  clustering,   5-9
  customize results,   5-24
  suggested content,   5-1
  top-N documents,   5-21
query configuration,   2-4
query-time authorization
  comparison with ACLs,   11-7
  configuration,   11-20

R

relevancy boosting,   2-3
  limitations,   12-22
result filter,   7-22, 13-48, 13-49
ResultFilterPlugin class,   11-21
ResultFilterPlugin interface
  API,   13-49
robots META tag,   4-7, 12-8
robots.txt file,   4-7, 12-8, 13-45
robots.txt protocol,   4-7, 12-8
rollover_key,   11-52
rules
  domain,   4-4
  path,   4-4

S

schedules,   2-3
  failed,   2-3
  understanding,   12-5
search attributes
  default,   4-25
search performance,   2-3
search server configuration,   12-21
searchctl command,   2-7
searchctl rollover_key,   11-52
searching
  basic search,   3-4
  date attributes,   3-9
  restricting,   3-12
  source groups,   3-4, 3-9
searchsys
  administrative user,   11-2, 11-12