

Oracle® Secure Enterprise Search

Readme

10g Release 1 (10.1.6) for Linux x86-64

B28691-02

April 2006

Oracle Secure Enterprise Search (SES) 10g Release 1 (10.1.6) Readme includes the following topics:

- [Features in this Release](#)
- [Version Information](#)
- [Removing Oracle Secure Enterprise Search Before Reinstalling](#)
- [Known Issues](#)
- [Documentation Errata](#)
- [Documentation Accessibility](#)

Features in this Release

Following are the features:

- Simple install
- Web-like query language
 - Simple search
 - Advanced search
 - Parametric search
- Relevance boosting
- Duplicate elimination
- Crawl and search
 - Web pages
 - Files
 - Databases
 - OracleAS Portal instances, both private and public pages
 - E-mail and mailing list (IMAP)
 - User-defined
- Web-based administration GUI
 - Administer crawl activities
 - Monitor crawl and query statistics

ORACLE®

- Secure search
 - SSO, http basic and form-based, cookies and https
 - Self service
 - ACL stamping
 - ACL crawling (via custom crawl)
 - Query time authentication
- Federated search
 - Federate to Oracle Secure Enterprise Search instances
- Search API via Web Services

Notes: The remote crawler feature is not available in this release.

For processing multiword queries, the default behavior in Oracle SES is "AND". This is a change from the earlier Beta default behavior of "OR".

Version Information

To identify detailed version information for your Oracle Secure Enterprise Search installation, do the following:

```
% $ORACLE_HOME/jdk/bin/jar -xvf $ORACLE_HOME/search/lib/search_query.jar  
"oracle/search/query/buildinfo.properties"
```

```
% cd ./oracle/search/query
```

The `buildinfo.properties` file contains detailed build information that can be provided upon request to Oracle Technical Support.

Removing Oracle Secure Enterprise Search Before Reinstalling

Ensure that earlier versions of Oracle SES have been completely removed before reinstalling it.

To remove Oracle SES, run `$ORACLE_HOME/install/deinstall_ses`. If the Oracle SES instance was bound to a port less than 1024 and the middle tier ran as root user, then after removal there could be files left under `$ORACLE_HOME`. After the removal completes, log in as the root user and delete those files.

Note: `$ORACLE_HOME` represents the directory where Oracle SES was installed.

Known Issues

For updated information on bugs, workarounds, and best practices, go to the following site:

OracleAS Portal Issues

Following are the issues:

- Secure OracleAS Portal crawling requires the latest version of OracleAS Portal.

Workaround: The following table lists patch information necessary for each version of OracleAS Portal for which Oracle SES certifies. You can apply the one-off patch on top of the listed version, or you can wait to upgrade to the next patchset indicated. The one-off patches are available on *OracleMetaLink* at

<https://metalink.oracle.com>

Enter the bug number for the appropriate OracleAS Portal release to get the patch.

Table 1 OracleAS Portal Patch Necessary for Secure OracleAS Portal Crawling

OracleAS Portal Release	Patchset to Upgrade	One-off Patch
9.0.4	9.0.4.3.0	This is tracked with bug 4637116. This should be applied on top of 9.0.4.2.0.
10.1.2	10.1.2.2.0	This is tracked with bug 5043725. This should be applied on top of 10.1.2.0.2.
10.1.4	10.1.4.1.0	This is tracked with bugs 4892619 and 5024386. This should be applied on top of 10.1.4.0.0.

Installation Issues

An exception error occurs if the **Back** button is clicked on the **Specify Inventory Directory and Credentials** page.

This is tracked with bug 5091607.

General Issues

Following are the general issues:

- The http response header from an Oracle SES server contains the SERVER version information that may be used to exploit potential vulnerabilities of the system.

Workaround: Override the SERVER field in http response header by performing the following steps:

1. Locate the file: `$ORACLE_HOME/bin/searchctl.pl`.
2. Add the following characters in bold:

```
sub startUS()
{
    print "\nStarting Search midtier .... please wait.\n\n";
    my $scmd = $PFX.$JAVAW.$params." -Dhttp.webdir.enable=false
-Doracle.oc4j.localhome=$FFB
-Djava.security.properties=".catfile($JH,'home','config','jazzn.secur
```

```
ity.props').' -Dhttp.server.header=unknown
-Doracle.security.jazn.config='.catfile($JH,$INST_
NAME,'config','jazn.xml').' -jar '.catfile($JH,'home','oc4j.jar').'
-userThreads -config '.catfile($JH,$INST_
NAME,'config','server.xml').' -out '.catfile($JH,$INST_
NAME,'log','oc4j.log').' -err '.catfile($JH,$INST_
NAME,'log','oc4j.log').$SFX;
```

3. Restart the middle tier with `searchctl restart`.

This is tracked with bug 5088104.

- Oracle SES does not support a default entity to attribute mapping mechanism for XML content.

Workaround: A custom crawler plug-in is required to crawl and index XML. (see the *Oracle Secure Enterprise Search Administrator's Guide* for details).

This is tracked with bug 4501161.

- The e-mail message browser in the default query application is unable to display attached HTML files for crawled e-mail messages that were originally sent from certain mail clients (for example, Mozilla Thunderbird). An `UnsupportedEncodingException` will be shown instead.

This is tracked by bug 5104659.

Administration Issues

Following are the administration issues:

- Table sources cannot handle quoted identifiers for the table or column names. For example, table or column names containing spaces, lowercase characters, or non-alphabetic characters other than underscore (_), dollar sign (\$), and pound sign (#) are not supported.

Workaround: Create a view using nonquoted identifiers on top of the table having quoted identifiers. Then crawl that view.

This is tracked with bug 5093512.

- If you are administering an Oracle SES instance with existing sources and schedules and you accidentally click **Recover** on the **Global Settings - Configuration Data Backup and Recovery** page, then you should see the error message "Recovery is only allowed on a freshly installed system. No sources or schedules should be created before recovery." At this point, triggers and foreign key constraints in the embedded database will have been erroneously disabled. This may cause sporadic problems if you continue to use this Oracle SES instance.

Workaround: Contact support or access Metalink for the support instructions to re-enable the triggers and constraints in your Oracle SES instance.

This is tracked with bug 5086309.

- Limitations in the range of characters that can be used in the Suggested Links feature: Suggested links does not support keywords with a hyphen, such as e-travel. The characters `+-!~` in a query string are ignored while matching suggested links, and the characters `~!@#%^(*)_+=[\{}|;:?,./<>'"&` should be ignored in suggested links.

This is tracked with bugs 4434993, 4280409, and 4280437.

- Detailed crawler statistics does not display host names for URLs that are not of the form: protocol://host/path.

This is tracked with bug 4247515.

- When an end user changes the credentials (user name/password) for a subscribed (self service) source after it has been crawled, the earlier documents crawled with the old credentials remain in the index. This may cause unexpected behavior for e-mail sources, in which the old documents may share the same URL with new e-mail documents from the most recent e-mail account. In this situations, some e-mail messages from the most recent crawl using the new credentials may not be indexed if an old message shares the same folder and message IDs.

Workaround: Update the crawler recrawl policy to **Process All Documents** on the **Home - Schedules - Edit Schedules** page, and recrawl the source.

This is tracked with bug 5048374.

- The administration tool states that the following file format types are supported:
 - dBase
 - Paradox
 - Framework

These three file format types are *not* supported.

This is tracked with bug 5102440.

Secure Search Issues

Following are the secure search issues:

- If the SSL port for the connection to an Oracle Internet Directory server was incorrectly specified in the **Global Settings - Directory Setup** page, then the connection operation will succeed, but ACL security group resolution for search queries will be disabled. For example, suppose user1 is a member of group1, and a group ACE is added to grant group1 access to a source. User1 cannot access documents in that source, because the user's group membership is not resolved.

Workaround: Reconnect to the Oracle Internet Directory server using the correct SSL port.

This is tracked with bug 5105083.

- If it is necessary to switch from the current Oracle Internet Directory server to a different one, or to change any aspect of the connection configuration (such as realm), then the ACL for all sources currently protected using Oracle SES ACL must be cleared (that is, all entries deleted) before registering the new Oracle Internet Directory server.

Workaround: The old ACLs should be manually deleted when switching from one Oracle Internet Directory server to another. Then the new ACLs added again.

This is tracked with bug 5048501.

- The documented restrictions about changing ACL policies are not enforced for mailing list sources.

Workaround: The ACL policy must be updated when the crawl is finished.
This is tracked with bug 5063158.

- The documented restrictions about changing ACL policies only work if the **Authorization** page for the source in the administration tool is current. If the page is not current, then you may be able to update the ACL policy in a way that should be disallowed.

Workaround: Before updating the ACL policy, ensure that the **Authorization** page is current by clicking the browser **Refresh** button.

This is tracked with bug 5065426.

- When Oracle SES is not connected to Oracle Internet Directory, the ACL policy can be set to **Oracle SES ACL** or **ACLs Controlled by Source**. This should not be allowed.

Workaround: Do not set the ACL policy to **Oracle SES ACL** or **ACLs Controlled by Source** when Oracle SES is not connected to Oracle Internet Directory. It should be set to **No ACL**.

This is tracked with bug 5064538.

- If you are writing a custom query application that communicates with an Oracle SES federated setup, and if a source on the Oracle SES slave instance is protected by query-time authorization, then any session context created on the Oracle SES master instance will not be propagated to the `QueryTimeFilter` on the slave instance. This is an issue when the custom query application is utilizing the `setSessionContext` Web Services operation to perform user authentication without an Oracle Internet Directory server.

Workaround: Use the standard Oracle Internet Directory user authentication setup for the Oracle SES master instance. In this configuration, the `AUTH_USER` value representing the Oracle Internet Directory authenticated user will be propagated to the `QueryTimeFilter` on the slave node.

This is tracked with bug 5069128.

- Private results may not be returned if you are protecting Oracle SES using SSO mode (2) (that is, only private content is protected by SSO) and also using either the Web services login method or federation.

Workaround: Remove the `filter` element in the `web.xml` file. Follow these steps:

1. Open the `web.xml` file found in `$ORACLE_HOME/oc4j/j2ee/oc4j_applications/applications/search_query/query/WEB-INF/web.xml`. You will see two elements `filter` and `filter-mapping`. Comment them out:

```
<!-- commenting filter and filter-mapping due to bug 5072567
<filter>
  <filter-name>RequestFilter</filter-name>
  <filter-class>oracle.search.query.RequestFilter</filter-class>
</filter>

<filter-mapping>
  <filter-name>RequestFilter</filter-name>
  <servlet-name>OracleSearch</servlet-name>
</filter-mapping>
```

-->

2. Restart the middle tier with `searchctl restart`.

This is tracked with bug 5072567.

- Cannot use the single sign-on (SSO) authentication for Web sources. SSO authentication looks for one signature, which is the name of the SSO login form. The name is currently hard-coded as "LoginForm", as it is the default form name for the SSO server.

Workaround: Make the form name a parameter in `eq$crawler_config_default`. When new form names need to be added, an `INSERT` statement like the following can be issued to add a new value into the global config table:

```
INSERT INTO EQ$CRAWLER_CONFIG_DEFAULT(CCD_CW_ID, CCD_PNAME, CCD_PVALUE)
values(-2, 'CC_SSO_FORM_NAME', 'NewFormName');
```

This is tracked with bug 4727588.

Search Term Issues

Following are the search term issues:

- Wildcard search inside a phrase is not supported.
This is tracked with bug 4451462.
- Boolean OR operator is not supported between advanced search operators. (for example, site & filetype query operators).
This is tracked with bug 4451445.
- Very long search terms cause an exception in the Web service query interface.
This is tracked with bug 4446793.
- Browse results do not include file sources.
This is tracked with bug 4285489.

Internationalization Issues

Following are the internationalization issues:

- The first character of highlighted text (KWIC) is not displayed for multibyte words.
This is tracked with bug 4311907.
- The administration environment language is reset to English after having been set to another language if not properly logged out.
This is tracked with bug 4424711.
- Browse categories render incorrectly in BIDI.
This is tracked with bug 4198113.
- If Oracle SES is installed in an operating system with European locale, then the crawler will finish with `Failed` state with the following error in the crawler log file:

ORA-00933: SQL command not properly ended.

The first time crawl that raises the error is successful; that is, all documents crawled are searchable. But recrawl of the same source will be blocked. After applying the workaround, you should be able to recrawl.

Workaround: Set the machine locale to English.

For example:

```
setenv LANG en_US
setenv LC_ALL en_US
searchctl stopall
searchctl startall
```

This is tracked with Oracle bug 5113622.

- If the browser locale is Portuguese, then search fails with an Oracle internal error. The locale code for Portuguese is a reserved word for Oracle Text, and the query introduces a syntax error.

Workaround: Specify a locale that is not Portuguese.

This is tracked with Oracle bug 5065141.

Crawler Plug-in Issues

Following are the crawler plug-in issues:

- If the crawler plug-in specifies a default parameter value that is encrypted, then the actual value is not propagated through the administration tool. That is, the encrypted value "*****" is saved. This occurs only on the **Home - Sources - Create User-Defined Source** page.

Workaround: With a custom crawler plug-in, do not provide a default value for a parameter that is encrypted. Instead, enter the encrypted value manually when creating the user-defined source. Otherwise, if a default value for an encrypted parameter is supplied, then manually re-enter the default value, or a new value, in the input box before creating the user-defined source.

This is tracked with bug 5076029.

- In the crawler plug-in interface, when a document is empty or is in pure text format (not HTML), document attributes may not be inserted correctly into the index. This may affect attribute search and search within a source group.

Workaround: Add a start tag and end tag in the document content; for example, `<html></html>` for empty or text documents before submitting the document to the crawler.

This is tracked with bug 5047114.

- When editing user-defined sources, the following pages in the administration tool do not apply to the crawler plug-in:
 - **Authentication** page
 - **Document Types** page
 - **Crawling Parameters** page: Only the following fields on this page can be set for user-defined sources: **Number of crawling threads**, **Enable language detection**, **Default language**

- When specifying an access control list principal to the `DocumentAcl.addPrincipal` method, the principal format cannot be in DN format.

Workaround: Specify principal in GUID format.

This is tracked with bug 5089586.

- The crawler could get an Out of Memory error as document content is read into memory if the plug-in submits text documents totalling around 100M.

This is tracked with bug 5043491.

- The crawler does not enforce maximum document size and mimetype checking for plug-in crawl.

Workaround: The plug-in must handle mimetype rejection and large document rejection itself.

This is tracked with bug 5043493.

Web Services API Issues

Following are the web services API issues:

- The `logUserClick` API is not supported.

This is tracked with bug 5090058.

- Deploying an application using the Web Services APIs in the same middle tier (OC4J container) as Oracle SES is not supported if the Oracle SES crawlers require HTTP proxy settings.

Workaround: Oracle SES HTTP proxy settings are applied to the entire OC4J container. Invocation of the WS APIs should be done without using a proxy. That is, this particular URL to the custom Web application should be excluded from the proxy settings. For this, make the settings in `searchctl.pl` file directly, so that it will be effective for the entire OC4J container.

1. Open `$ORACLE_HOME/bin/searchctl.pl`.
2. Find the `startUS` subroutine.
3. Add the following options to the `$cmd` variable after the first doublequote:

```
-Dhttps.proxyHost=<host> -Dhttps.proxyPort=<port>
-Dhttp.proxyHost=<host> -Dhttp.proxyPort=<port>
-Dhttp.nonProxyHosts='<*.domain>'
```

where `<*.domain>` is the domain you want to exclude from the proxy settings

4. Restart the middle tier with `searchctl restart`.

This is tracked with bug 5102814.

Federated Search Issues

Following are the federated search issues:

- After the federator is deployed, Oracle Internet Directory connection and disconnection cannot be done with the Oracle Secure Enterprise Search administration tool.

Workaround: Connect to, and disconnect from, Oracle Internet Directory through a command line tool.

Enter the following to connect to Oracle Internet Directory from the command prompt:

```
$ORACLE_HOME/bin/searchctl config action=setup_oid oid="<oid host>"
oid_port=<oid port> oid_sslport=<oid SSL port> oid_subscriber_dn="<realm
distinguished name>" oid_user_dn="<administrator user name>"
oid_passwd=<administrator password> jdbc_str="<jdbc connect string>"
dba_user=eqsys dba_passwd=<SES password>
```

The Oracle Internet Directory host, Oracle Internet Directory port, Oracle Internet Directory SSL port, realm distinguished name, administrator user name, and Oracle Internet Directory password should be entered exactly as they would be entered on the **Global Settings - Directory Setup Page** in the Oracle SES administration tool. The Oracle Internet Directory host, realm distinguished name, and administrator user name should be enclosed in quotes (as shown).

The JDBC connect string can be obtained from the `$ORACLE_HOME/search/webapp/config/search.properties` file by examining the `connection.url` entry. The entry should have the form `connection.url=jdbc:oracle:thin:@<SES host>:<DB port>:<DB SID>` For the `searchctl` command to connect to Oracle Internet Directory, only the host, port, and SID components should be used (that is, everything following, but not including, the '@' character).

Enter the following to disconnect from Oracle Internet Directory from the command prompt:

```
$ORACLE_HOME/bin/searchctl config action=disconnect_oid jdbc_str="<jdbc
connect string>" dba_user=eqsys dba_passwd=<Oracle Secure Enterprise
Search password>
```

This is tracked with bug 4962145.

Documentation Errata

This section corrects errors in the existing documentation.

Online Help

The **Online Help** page for the end user search states the following:

A query can have only one filetype shortcut. The following filetype extensions are supported: doc, html, pdf, txt, rtf, ppt, ps, xls.

Postscript is not supported, so ps should be ignored.

Documentation Accessibility

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

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

Accessibility of Code Examples in Documentation

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

Accessibility of Links to External Web Sites in Documentation

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

TTY Access to Oracle Support Services

Oracle provides dedicated Text Telephone (TTY) access to Oracle Support Services within the United States of America 24 hours a day, seven days a week. For TTY support, call 800.446.2398.

Oracle Secure Enterprise Search Readme, 10g Release 1 (10.1.6) for Linux x86-64
B28691-02

Copyright © 2006, Oracle. All rights reserved.

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

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

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

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

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

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

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

