New Query Features in Secure Enterprise Search 10.1.8.2

An Oracle White Paper
October 2007

New Query Features in Secure Enterprise Search 10.1.8.2

Executive Overview

Oracle Secure Enterprise Search version 10.1.8.2 introduces a new advanced query syntax for power users. This document looks at the capabilities of the new query syntax and the setup (where necessary) in order to use it.

This document is intended to complement, but not replace, the *Oracle Secure Enterprise Search Administrator's Guide.* Readers of this document are strongly advised to consult the *Administrator's Guide* after learning about a new query feature through this document.

Introduction

Previous versions of Secure Enterprise Search (SES) have had only two choices for querying: a simple "basic search" and a separate "advanced search" page. Basic Search had a few options, such as phrases and excluded terms. Advanced Search had the capability to include attribute searches. In 10.1.8.2, the syntax for basic search is considerably enhanced to allow for much more advanced searching.

Whenever possible, the advanced search has been designed not to get in the way of novice users. For this reason, syntax is not rigidly enforced. If a query contains syntax errors (as is common if text containing punctuation is pasted into the search box) then SES will transform this into a "best guess" query.

Recap – 10.1.8.1 Syntax

First we'll take a look at the limited syntax available in previous releases. All these still work in 10.1.8.2. In this section and the rest of the document, we will use square brackets (box brackets) to quote the text of the query and/or the matched term. For example: [the quick "brown fox"]. This is to avoid confusion with double-quotes ([""]) which have special meaning within SES syntax.

Single Word Search

A single word will match all documents that contain that word.

Multiple Word Search

Two or more words will be treated as an AND search. A search for [green eyes] would match all documents containing the word [green] and the word [eyes].

Compulsory Exclusion

Prefixing a word with the minus sign ([-]) would exclude any documents containing that word from the search. So a search for [green -eyes] would match all documents containing the word [green] but not the word [eyes].

Phrase Search

A phrase search is any set of words enclosed in quotes. For example the search [“green eyes”] (with the quotes) would search for documents which contain the word [green] followed immediately by [eyes].

Note: There is also a “compulsory inclusion” operator ([+]). However this normally has no effect on a search since all words are included anyway. This operator only has an effect if an undocumented switch is used to provide “default OR” searching rather than “default AND”.

New Syntax in 10.1.8.2

Operators affecting multiple terms

Phrase Searching [“”]

As before, any set of words within quotes is a phrase search.

AND Operator (default)

Also as before, any two or more terms (not quoted) are automatically treated as an AND search.

OR Operator [|]

The new operator [|] (vertical bar, or pipe symbol) is an OR conjunction. It can be used between two terms or sets of terms. So for example the search [dog | “male canine” | great dane] will find documents containing BOTH the words [great] and [dane] OR the phrase [male canine] or the word [dog]. It can be seen from this analysis that AND binds more tightly than OR.

Proximity [“…”~]

The proximity operator tilde must be used in conjunction with quotes. It may be used with or without a spanning distance. For example [“search secure”~10] will find any document with the terms “search” and “secure” with no more than 10 words between them (in any order). The maximum spanning distance allowed is 100, and this is also the default if not specified. If more than two terms are specified, the spanning distance is the maximum number of non-matching terms which occur between the matching terms. For example

Text: [the **quick** brown **fox** jumps over the lazy **dog**]

Search: [“fox dog quick”~] matches. The order is not important, and the default spanning distance of 100 will be used.

Search: [“fox dog quick”~5] matches, since there are 5 non-matching terms.

Search: [“fox dog quick”~4] does *not* match – there are too many terms between.

Grouping [()]

You can override the default precedence of operators by putting parentheses round expressions. For example [(blue | green) eyes] would match documents containing [blue] and [eyes] and also documents containing [green] and [eyes].

Term Expansion Operators

Wildcards – Single [_] and Multiple [*]

The special character underscore [_] represents any single letter. So [fus_on] would match [fusion] or [fuseon]. The star or asterisk [*] matches any sequence of zero, one or more characters – so [fu*n] would match [fun], [fuon], [futon], [fusion], [function], etc.

There are some limitations with wildcards. You cannot use wildcard characters at the beginning of words – [_unction] would not be allowed. Nor can you use a wildcard which would result in too many expansions – [a%] is unlikely to work. The number of expansions that would result from any particular search term would depend on the data indexed. The limit in the current version is 5000 expansions.

Fuzzy Match [~]

The tilde operator [~] used after a word indicates that the word should be expanded to include all words, which are spelled or sound similar. For example [orakle]~ would match [oracle] as well as [orakle]

Do not confuse a tilde used after a single word with a tilde used after a phrase: the NEAR operator covered above. Also do not confuse it with a tilde used *before* a word, which is the synonym operator, covered below.

Synonym [~]

The tilde operator used before a term indicates that the system should search for that term and any synonyms of it. Synonyms are words with the same meaning, and are defined by the administrator by the process of thesaurus loading described in the *Secure Enterprise Search Administrator's Guide.*

For example, if the administrator had defined “cat”, “feline” and “moggy” as synonyms in the thesaurus, then [~cat] would match any of the three words.

Synonyms can be phrases as well as single words, in which case the phrase should be enclosed in double quotes. [~”United States”] might expand to include [USA].

Do not confuse tilde at the beginning of words or phrases with tilde at the end of words: (the fuzzy match operator covered above) or tilde at the end of phrases (the NEAR operator).

Narrower terms [<]

Another thesaurus operator, "left angle bracket" [<] used before a word or phrase expands the search to include all narrower (more specific) terms in a thesaurus hierarchy as defined by the administrator. So [<"United States"] might expand to include [Alabama], [Alaska], [California]…

Broader terms [>]

"Right angle bracket" [>] used before a word or phrase expands the search to all broader terms in a thesaurus hierarchy. So [>"San Francisco"] might expand to include [California] and [United States]. Note that broader and narrower terms are symmetric – if [California] is defined as a broader term of [San Francisco], then [San Francisco] is automatically a narrower term of [California].

For information on how to load thesaurus information for use as synonyms, broader terms and narrower terms, please see the *Oracle Secure Enterprise Search Administrator's Guide.*

Attribute-limiting Operators

You can use the [:] (colon) operator to restrict your search to documents with particular attribute values. As a simple example, [title:oracle] restricts the search to documents with the word [oracle] in the title. Since attributes may be of various types (string, number or date) there are different operators to deal with each type.

Exact match [=]

The exact match operator may be used with all attribute types. If more than one word is used, the phrase must be surrounded by quotes. For example: [title="oracle text"] will match documents with [oracle text] (or [Oracle Text] – it is not case-sensitive) as the *whole* title.

Contains match [^]

This operator only applies to string attributes, and is the default so will rarely be needed. It can, however, be included for clarity. A search for [title:oracle] is equivalent to [title:^oracle] and will find all documents with the word oracle somewhere in the title. The match will only be for whole words unless you use the wildcard [*]. You can use multiple words within a contains match by surrounding them with double quotes or parentheses [title:"kansas city" or [title:(kansas city)]. Note these are treated as phrases, not as an AND search as they would be if you used two words together outside of an attribute search.

Range searches

For numeric or date attributes, you can use [<], [>], [<=], or [>=], which each have their expected meanings (less-than, greater-than, less-than-or-equal-to, greater-than-or-equal-to).

Thesaurus within attribute search

For string attributes, the operators [<] and [>] have special meanings: narrower terms and broader terms respectively. See the later section on thesaurus searching.

Identifying Attributes

When doing an attribute-based search, you must of course know the name of the attribute you wish to include in the search. Full attribute names in SES can include spaces and special (punctuation) characters. In order to use such as attribute name in a search, you must surround the name with double quotes. Case is not significant. Thus, if we have an attribute called "Document Reviewed By" an attribute search using this might look like ["document reviewed by":"roger ford"].

Built-in Attributes

In general, in Secure Enterprise Search, there is no limit on the number of possible attributes. Any data source can provide arbitrary attributes. However, there are a number of built-in, or "special" attributes. The following table lists these:

Attribute Name	Description	Type	Usage Example
site	Host part of URL	String	site:www.oracle.com
sg	Source Group	String	sg:documentation
filetype	Abbreviated mime type	String	filetype:pdf

Notes:

1. Unlike normal attribute names, these special attribute names are case-sensitive. They must be entered in lower case. [fileType:pdf] is *not* the same as [filetype:pdf]. If you enter [fileType:pdf] as a search, SES will ignore the colon and look for documents which contain the words [filetype] and [pdf]. This is in line with the general policy of not rejecting any search because it fails to match a tightly-defined syntax.

2. The [site:] attribute search matches the *whole* of the server part of a URL. [site:oracle.com] does *not* match [http://www.oracle.com/foo].

3. The source group name for [sg:] must be specified in full. Case is not significant. If the source group name contains spaces or punctuation, it should be enclosed in quotes. For example if you have a source group called "Jim's Documents", you might use [sg:"jim's documents"] to reference it.

Oracle Text Pass-through

Oracle Text provides the underlying search capabilities for Secure Enterprise Search. Oracle Text has its own query syntax, which is intended more for use by applications or information professionals rather than casual end-users. However, there may be some circumstances under (such as running SES queries from an application that generates its own queries, where Oracle Text syntax is preferred. A query which is preceded by the special marker [otext::] will be passed directly to the Oracle Text engine without further processing by SES. The full syntax of an Oracle Text query is beyond the scope of this document. Please see the *Oracle Secure Enterprise Search Administrator's Guide* and the *Text Reference Guide* for further details.

Conclusion

Secure Enterprise Search 10.1.8.2 provides advanced query capabilities for the experienced user, or information professional, without compromising the ease of use of a simple "anything goes" query box.

Secure Search with Oracle Secure Enterprise Search
October 2007
Author: Roger Ford
Contributing Authors: Stefan Buchta, Jinyu Wang

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
www.oracle.com

This Document Is For Informational Purposes Only And May Not Be Incorporated Into A Contract or Agreement.