opentext™

SQL Text Indexing in Content Manager

White Paper

Contents

1 SQL Text Indexing in Content Manager

1.1 Overview

The SQL (Structured Query Language) Text Index feature allows organizations to index the Content Manager word indexes, such as Titles, Notes, text type Additional Fields, Classifications, Location Addresses, and so on, using native database indexing.

SQL Text Indexing is supported for MS SQL Server, Oracle, and PostgreSQL RDBMS and depends on the respective text indexing component of the RDBMS.

A pre-requisite for using SQL Text Indexing on MS SQL Server and Oracle is that you must have their associated 'full text' features installed and configured before upgrading or enabling SQL Text Indexing. PostgreSQL has built-in support for text searching, so no additional components are required.

For more information, see the link below from Microsoft:

- https://learn.microsoft.com/en-us/sql/relational-databases/search/

1.2 Why SQL Text Indexing

In 10.1, the Content Manager text indexing engine, which has been in place for many years, is fully replaced with the SQL text indexing capabilities built into modern relational databases.

This capability was first introduced in Content Manager 10.0 as an alternative. From 10.1 onwards, SQL text indexing is the only way to index metadata.

Note that, when upgrading from a version of Content Manager 10.0 or earlier, you will need to factor in a complete text reindex into your upgrade plan. Text search methods such as Title, Notes, and Any Word searches will not be available until the SQL Text Index is created and indexed.

The two immediate advantages of SQL Text Indexing are:

- Huge performance improvement when searching for records, sorting records, grouping records, or maintaining a unique column.
- Flexibility to search records with titles that include an apostrophe (') as part of the word, and/or a comma (,) as part of the phrase.

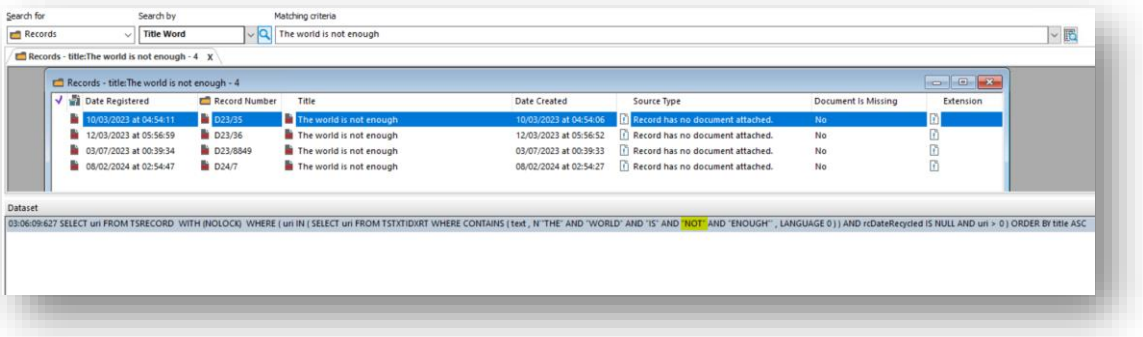
This change requires a re-acquaintance of search parameters.

1.2.1 Title Word

This search considers all the words used as search terms in text search.

EXAMPLE:

For searching records, you can search with matching criteria as “The world is not enough”, In this example, the word “not**”** is treated as a string. It will find records where the record title contains “The”, “world”, “is”, “not” and “enough”.



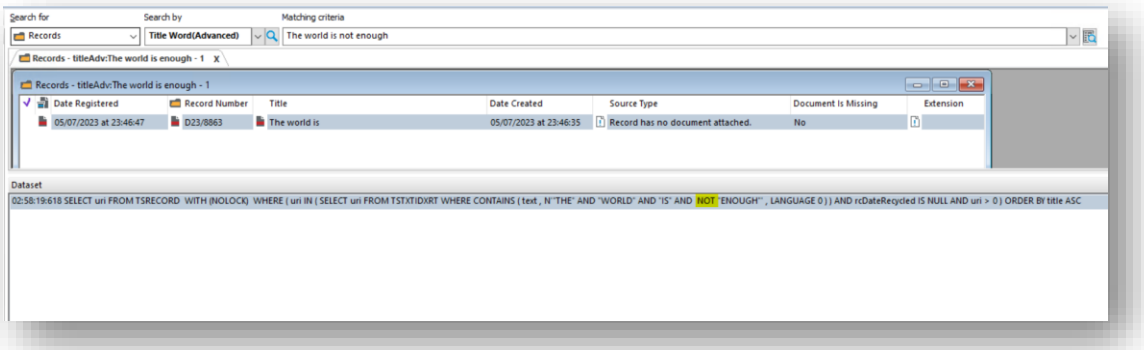
1.2.2 Title Word (Advanced)

You can search for records with titles that include an apostrophe (') as part of the word, and/or a comma (,) as part of the phrase. For example, what's happened, has happened! The Title Word (Advanced) search treats AND, OR, NOT, NEAR, and parentheses as keywords except if they are enclosed within the double quotes. The full-text query is formed by treating those keywords as logic operators.

*NOTE: If you are using logical operators (E.g., AND, OR, NOT, or NEAR) in the search criteria, it is **recommended** to use Advanced word searching*.

EXAMPLE:

For searching records, you can search with matching criteria such as "The world is not enough". In this example, the word **not** is treated as a logical operator. The search will find records where the record title contains "The", "world" and, "is" and does not contain "enough".



1.2.3 Advanced Word Search

NOTE: This feature is available only from Content Manager 23.3.

This option is only available for all word searches and available for selection only in the binary search editor. When selected, the search parser decodes and identifies all the logic keywords, e.g., **OR**, **AND**, **NEAR**, and **NOT**, and uses them as logic operators in text search, otherwise, all the words are used as search terms in text search.

1.3 Working with SQL Text Indexing

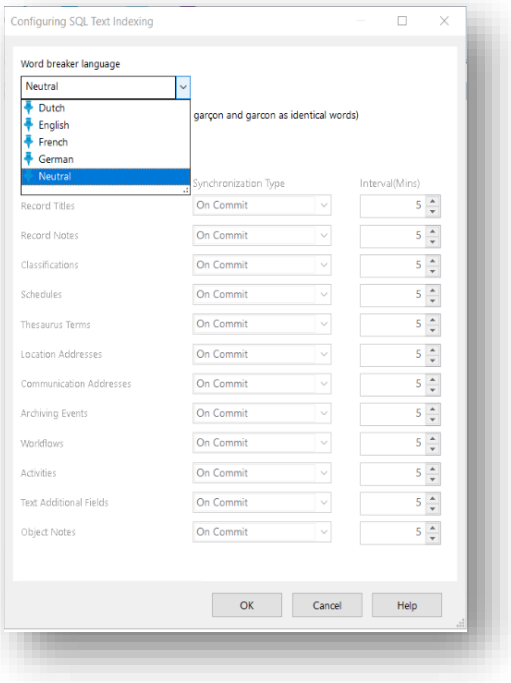
1.3.1 Accessing the SQL Text Indexing

To access the SQL Text Indexing options:

1. Expand **Datasets** in Content Manager Enterprise Studio.
2. Right-click the dataset you want to work with and point to **SQL Text Indexing**. The available options are:
    - Recreate
    - Check
    - Reindex
    - Synchronize Stoplist
    - Recreate indexes.

1.3.2 Configuring the SQL Text Indexing

1. Expand **Datasets**.
2. Right-click the dataset you want to work with point to SQL Text Indexing and click **Recreate**. The Configuring SQL Text Indexing dialog is displayed.



NOTE: The SQL Text Indexing dialog is also displayed as a part of the new dataset creation process; as well as when the SQL Text Index is re-created via the right-click SQL Text Indexing - Create option.

a) **Word breaker language** - by default this is set to **Neutral**. Select the word breaker language from the drop-down. The available languages are **Dutch**, **English**, **French**, and **German**.

b) If you wish to ignore the accents on characters so all words are indexed as the same words, select **Ignore accents on characters (treats garçon and garcon as identical words).**

c) For Oracle datasets only, the options Parallel Degree and the Word Type - Synchronization Types are available.

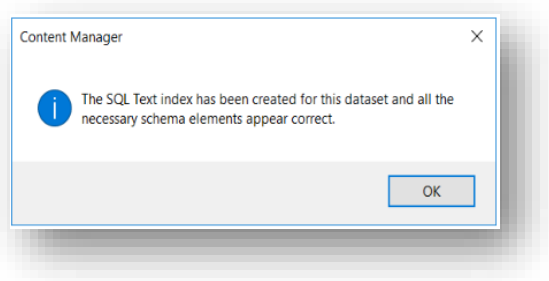
- **Parallel Degree** - by default this is set to 4 - set the number of parallel execution servers associated with a single operation. This can be set up to 99.
- **Word Type Synchronization Types** - from the drop-down list, select the **Synchronization Type**, select from:
  - **On Commit** - select this option to synchronize the indexes at the point of creation of the object.
  - **On Regular interval** - select this option to synchronize the indexes at regular intervals. The interval period is set in the **Interval (Mins)** option.

3. Click **OK**.

NOTE: When upgrading to SQL Text indexing it is necessary to run a Reindex to ensure all existing supported records/objects can be searched for using text search methods.

1.3.3 Checking the SQL Text Indexing

1. Expand **Datasets**.
2. Right-click the dataset you want to work with and point to **SQL Text Indexing** and click **Check**.



Content Manager will run a check to determine that all the required schema elements are available. If issues are found, a warning to repair the index will be displayed.

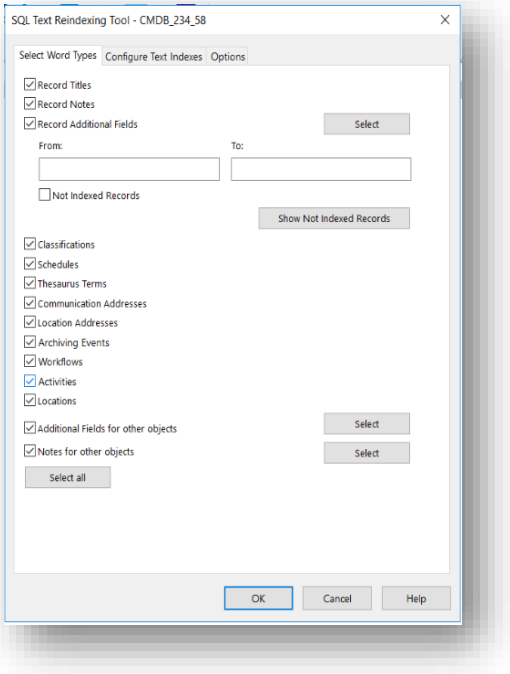
NOTE: Content Manager will display SQL Text Index is partially created against the Dataset name on the Enterprise Studio home page. This will indicate that a Check and repair is required. Once the repair has been completed, a re-index is required.

1.3.4 Reindexing the SQL Text Indexing

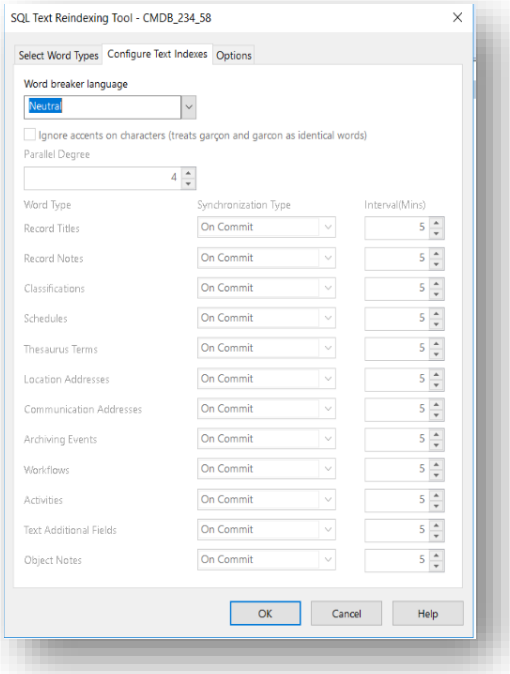
When a new SQL Text Index is created, or there has been ingestion of objects into the Content Manager dataset, it is necessary to reindex the dataset, so the objects are searchable.

1. Expand **Datasets**.

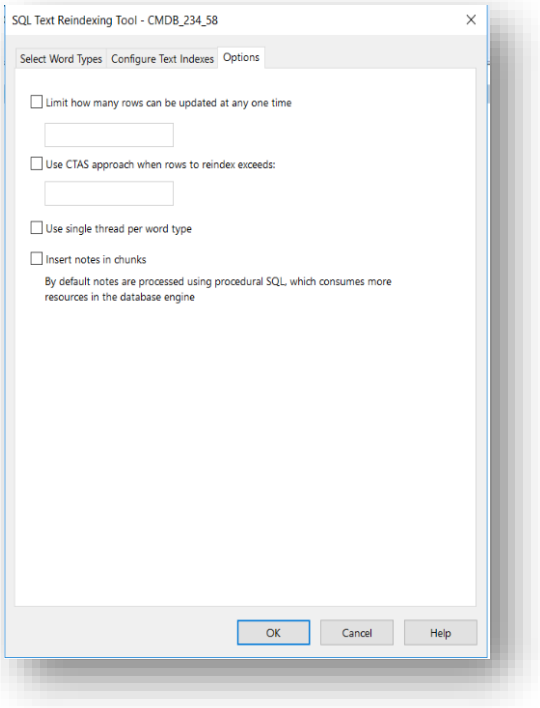
2. Right-click the dataset you want to work with and point to SQL Text Indexing and click **Reindex**. The SQL Text Reindexing Tool dialog is displayed.
3. In the **Select Word Types** tab, select the objects to be reindexed.



4. In the **Configuring SQL Text Indexing** tab, select the required options.

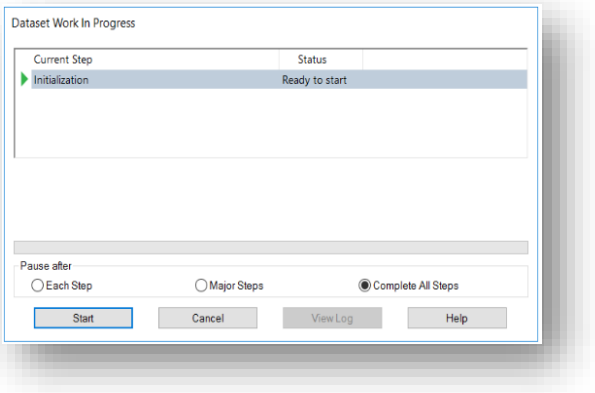


5. In the **Options** tab, select the required options:

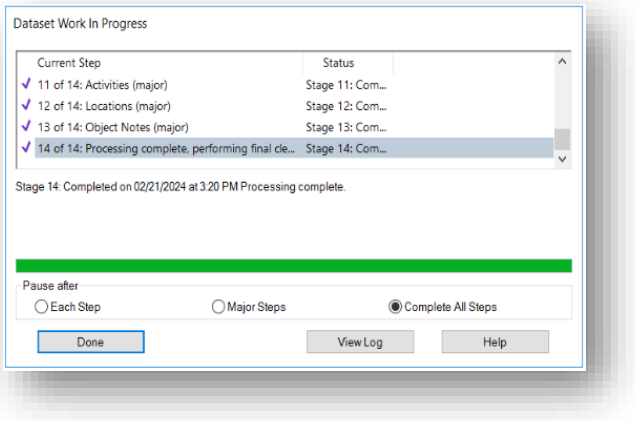


- **Limit how many rows can be updated at any one time** - select this option to limit the number of rows that are processed at one time and execute the operation in batches. This option may be of use when reindexing large datasets. Set the maximum number of rows to update at any one time. If the reindexing processes are timing out, reduce this number.
- **Use CTAS approach when rows to reindex exceeds** - option to reindex using CTAS approach that has better performance than the default one.
- **Use single thread per word type** - if checked, the re-indexing process runs in parallel using a single thread per word type.
- **Insert notes in chunks** - if checked, the notes and text of additional fields are copied using the Insert or Update command in chunks.

6. Click **OK** to run the reindex. The **Dataset Work in Progress** dialog will be displayed.



7. Select the **Pause After** option you want to apply:
    - **Each Step** - pause after each step is completed.
    - **Major Steps** - pause after each major step is completed.
    - **Complete All Steps** - complete all steps without pausing.
8. Click **Start** to start the upgrade process.
9. Once all steps are completed, click **Done** to close the **Dataset Work in Progress** dialog. If required, click **View Log** to view the generated log file.



1.3.5 Synchronizing the Stoplist

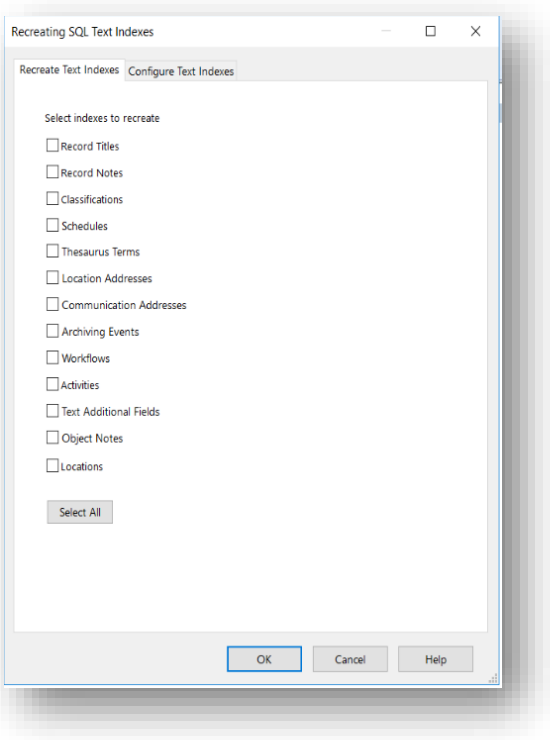
Select this option to synchronize the SQL Text custom stoplist with the Content Manager noise words. (For information about noise words, see section **NOISE WORDS**)

NOTE: If using PostgreSQL, the trimstopword PostgreSQL stop word file must be located in the PostgreSQL\share\tsearch_data folder before running the Stoplist synchronization.

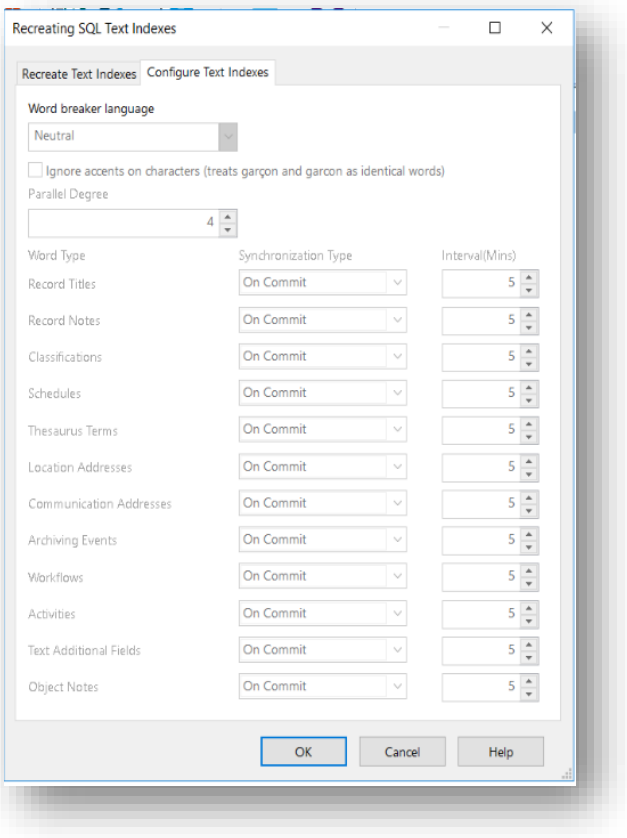
1.3.6 Recreating the SQL Text indexing

Select this option to recreate SQL Text indexes.

1. Expand **Datasets**.
2. Right-click the dataset you want to work with and point to SQL Text Indexing and click **Recreate indexes**. The Recreating SQL Text indexes dialog is displayed.



3. Select one or more checkboxes or click **Select All** to select indexes to recreate. The following are the options:
    - Record Titles
    - Records Notes
    - Classifications
    - Schedules
    - Thesaurus Terms
    - Location Addresses
    - Communication Addresses
    - Archiving Events
    - Workflows
    - Activities
    - Text Additional Fields
4. In the **Configuring SQL Text Indexing** tab, select the required options.



5. Click **OK**.

1.4 Limitations of using SQL Text Indexing in Content Manager

In SQL Text Indexing, the text/string (E.g., AND, OR, NOT, and NEAR) are treated as logical operators. For the existing Content Manager users, this was a major behavioral change. These reserved SQL keywords are used as pure strings, resulting in a mismatch of expected records.

To mitigate this, we introduced a new search option "Title Word (Advanced)" that honors the reserved keywords in SQL. To retain the user experience, we changed the "Title Word" search method to treat the reserved SQL keywords as pure strings.

2 Search in Content Manager – Recap

2.1 Overview

You can search for Content Manager items using practically all the information about them that is saved in Content Manager. There are search methods that are quick and simple and others that may

take a little longer to set up but enable you to combine criteria in a more sophisticated manner. The information below helps how to search in Content Manager.

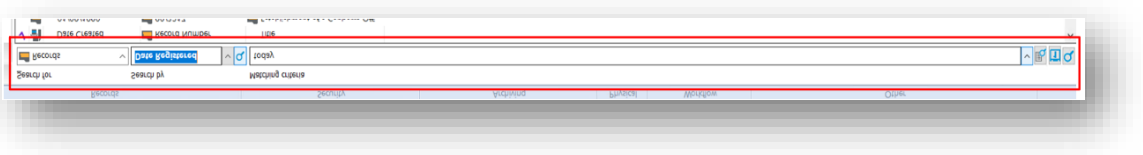
2.2 Quick search using just one criterion: Search pane

In the main Content Manager window under the toolbars, use the fields **Search For**, **Search By**, **Matching Criteria**, and click the **Run the search** button.

2.3 Search pane

Use the **Search** pane to perform fast and simple searches.

By default, it is visible and appears in your search toolbar at the top of your Content Manager window.



You can hide it or show it if it is not visible using the **View** menu command **Search Pane**.

The **Search** pane works much like a search using the **Search** menu; however, you can only search by one field.

The **Search** pane has these fields and buttons:

- **Search For** - to select the type of Content Manager item you are looking for, for example, records or Locations.

  It depends on your permissions and which type of objects you can search for.

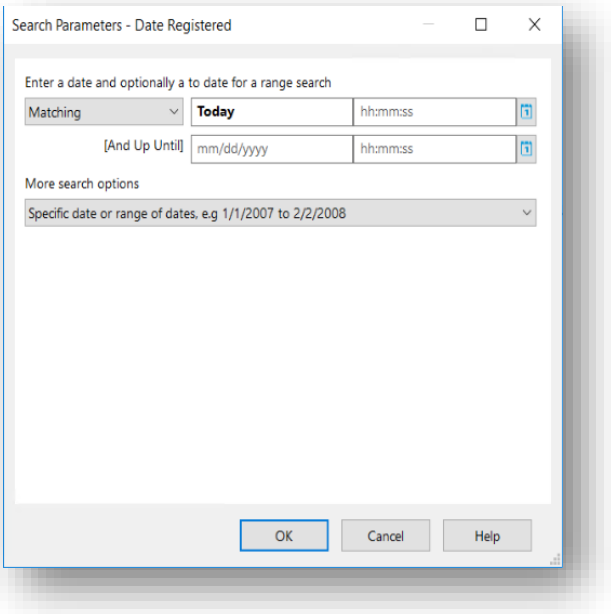
- **Search By** - to select the data or field to search by, for example, **Document Content** for document content, or **Date Created**.

  You can type the caption or field name, or the search method in the field.
  Click the **Down** arrow to the right of the field for the most recent data you searched by, favorites, and all terms alphabetically sorted. Separators are added to differentiate between recent data, favorites, and all terms.

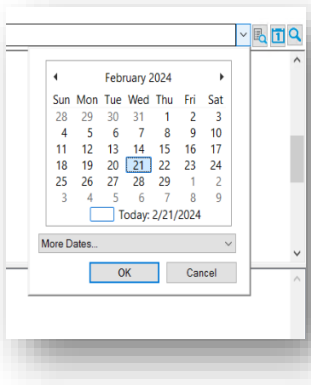
  Click the search button to the right of the arrow for the full list of methods you can use to search Content Manager in the **Search Methods** dialog box.

- **Matching criteria** - to enter the data to find, like a word, a date, or a Location.

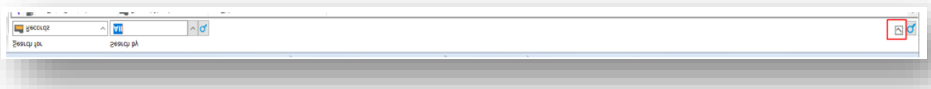
- Buttons to the right of the **Matching criteria** field:
  - **More options for entering the search value** button - displays the **Search Parameters** dialog box for more search options.



- **Select date from calendar** button - available when you select a date field in the **Search By** field - to select a date from a calendar, rather than typing it.



- **Check box** - available when you select a Boolean field in the **Search By** field - selected indicates YES and unselected indicates NO for the Boolean operator.



- **Run the search** button - click to execute the search and display the result.

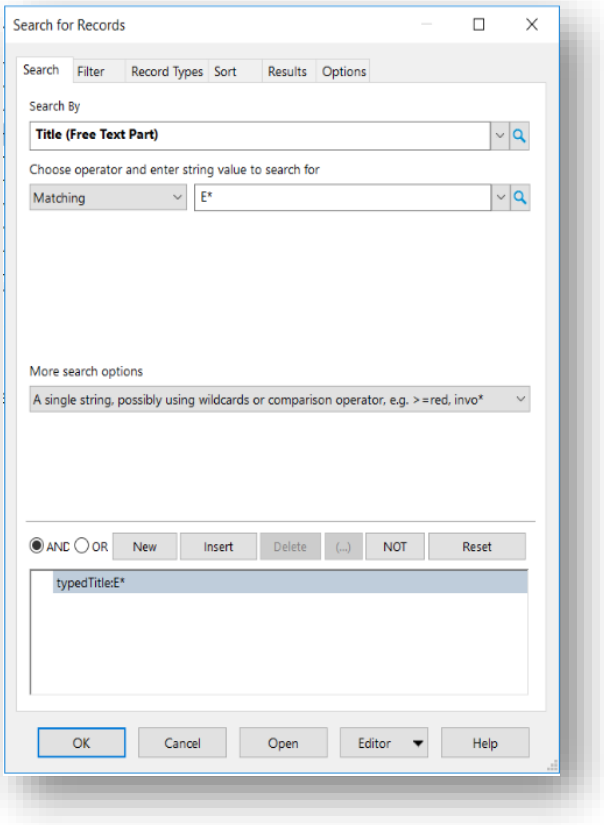
2.4 Easy search using one or more criteria: Search for ... dialog boxes

Use the **Search** pane to perform fast and simple searches.

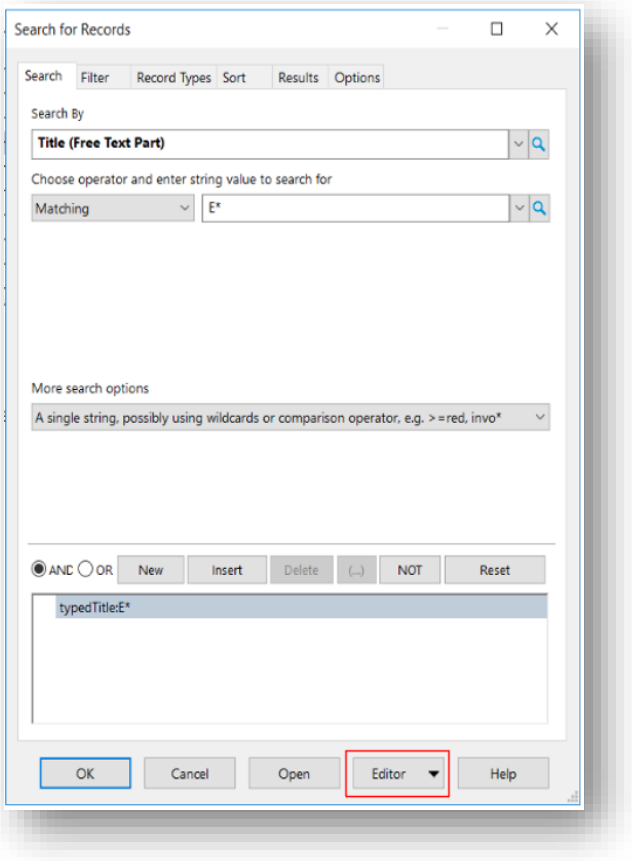
2.4.1 Search for ... dialog boxes

The **Search for ...** dialog boxes appear when you search for items in Content Manager.

This may be when you use the **Search** menu on the toolbar, or when you right-click a list and on the **Search** menu, click **Refine Search**, for example.

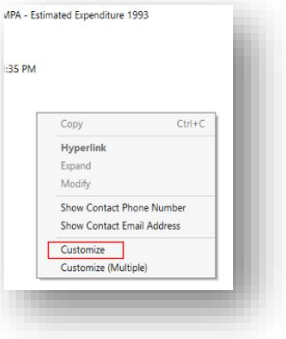


The **Search for ...** dialog box layout and fields depend on the type of Content Manager item you are looking for, and the search editor you are using. You can quickly change the search editor you are using with the button **Editor** at the bottom.



To sort the search results in a particular order and to customize the columns for the search results, use the **Sort**, **Filter**, and **Results** tabs and choose how to present the list.

Use the view pane's right-click menu and select **Customize** to configure Content Manager so that the information you want to see for each item is listed.



If necessary, to refine your search, right-click in the list of search results, and on the **Search** menu, click **Refine Search**.

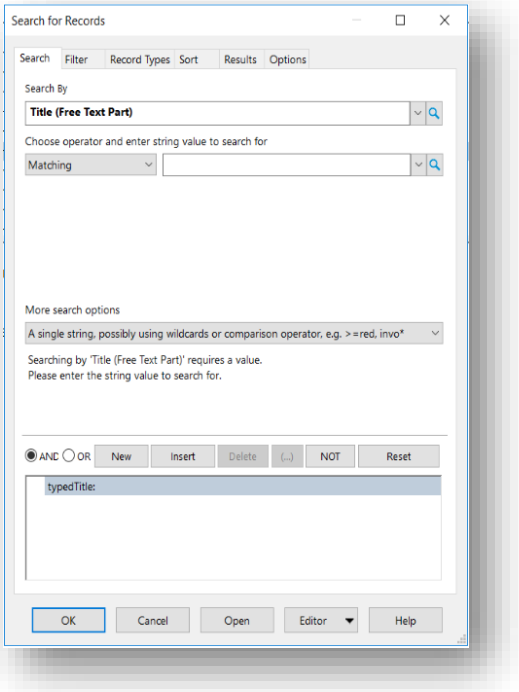
2.5 Advanced search

Using the string search editor, you can use the full capabilities of Content Manager searching, combining potentially large numbers of criteria, and using advanced filter and sort options.

The string-based search editor enables you to use search strings to take full advantage of Content Manager's advanced search capabilities.

You can use it to search for any item in Content Manager that you can search for.

1. On the **Search** menu, click what you are searching for, for example, **Find Records** to find records or **Find Locations** to find Locations.

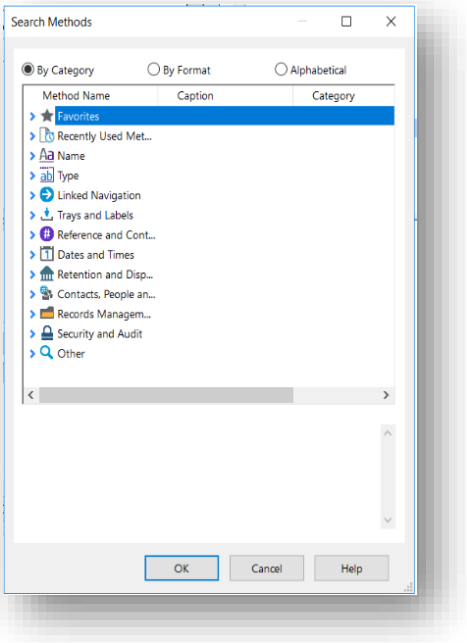


2. If it is not set, use the button **Editor** at the bottom and select **String** to use the string search editor.
3. In the **Search** tab, use the field **Enter search string** to build your search.

   You can type your search clauses manually or use Kwik Select to select them and string them together.

   a) Click Kwik Select and the **Search Methods** dialog box appears.

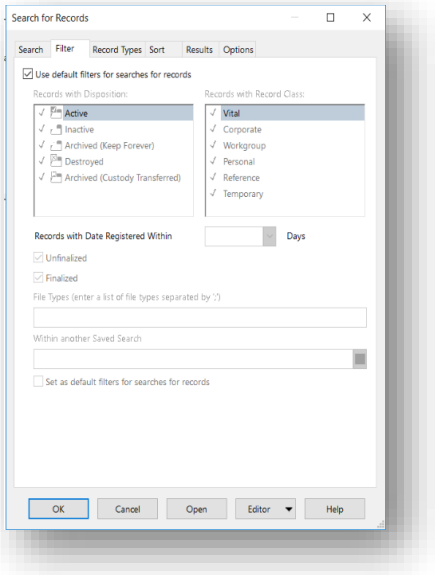
      - **By Category** - displays the search methods in categories
      - **By Format** - displays the search methods by their data format
      - **Alphabetical** - displays the search methods in alphabetical order.



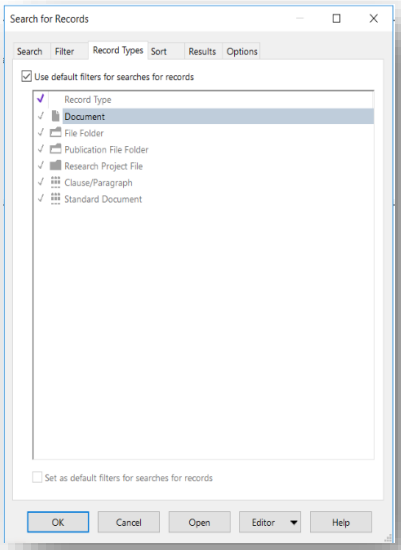
Select a search method and review the text below the list to find out more about it.

You can use the tabs to define your search further:

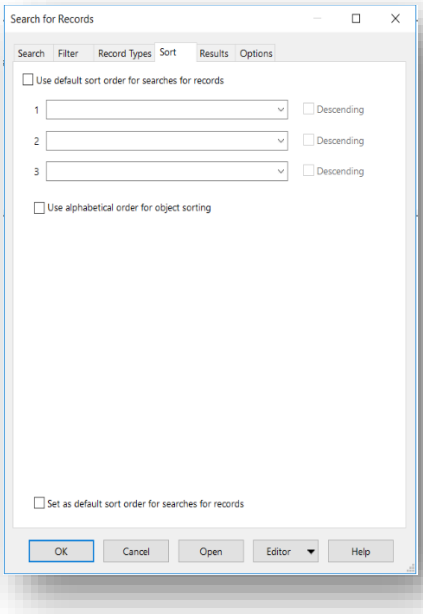
- **Filter** tab - Use the Filter tab on a Search dialog box to filter the search results.



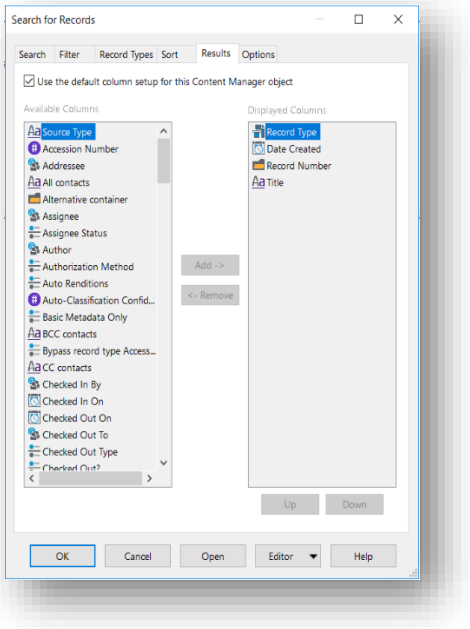
- **Record Types** tab - When searching for records, you can use the Record Types tab to filter your search results by the Record Types the records are using.



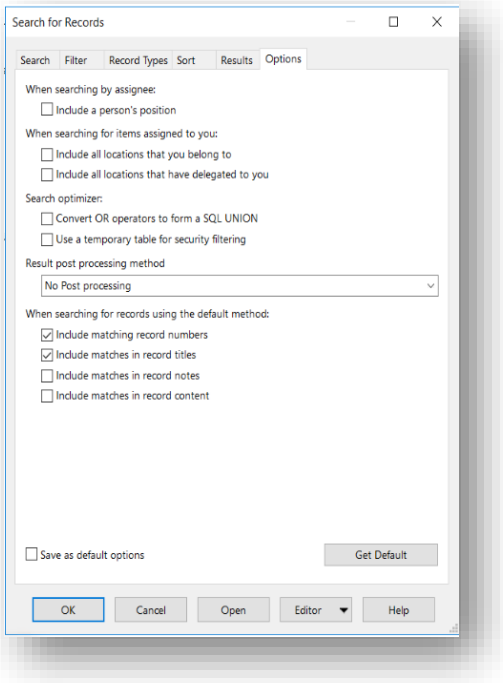
- **Sort** tab - In the Sort tab on a Search dialog box, use the following options to sort a search result list.



- **Results** tab - You can use it to choose the columns that Content Manager should use when displaying search results.



- **Options** tab - When you enter your criteria for a search, one of the tabs available to you is Options.



4. Click **OK** and Content Manager displays the search result.

3 Noise Words

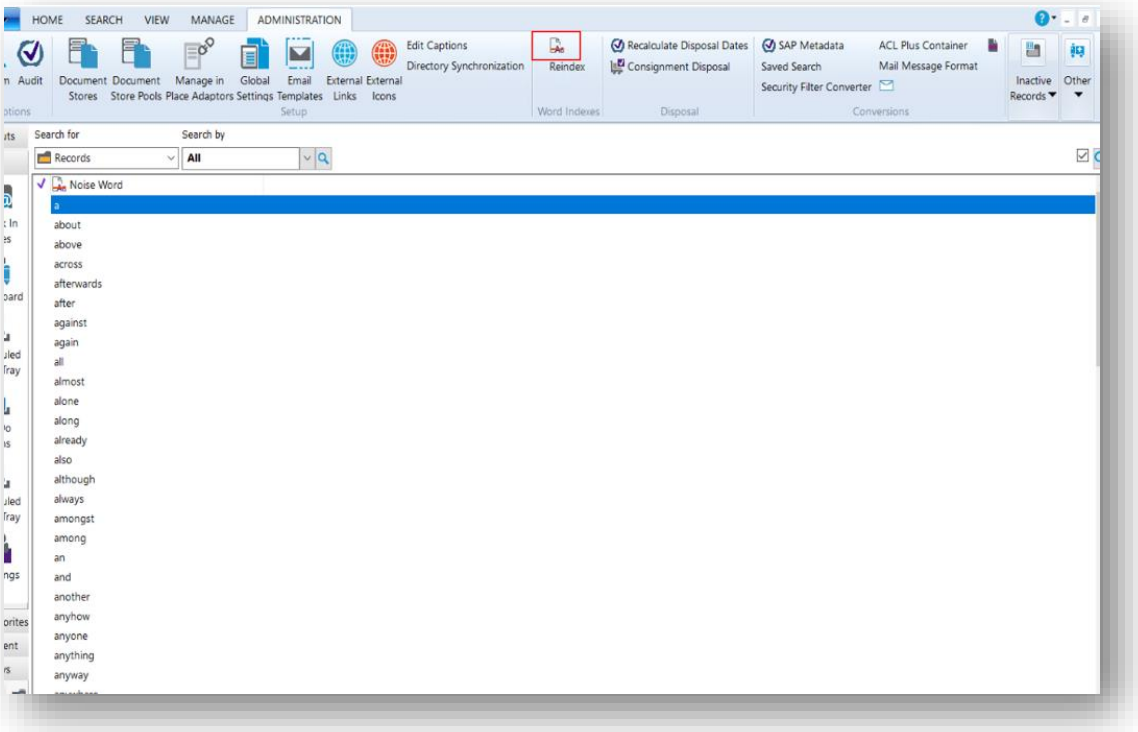
This is a list of words that will not be indexed if they are used in record titles, notes, Thesaurus terms, and Categories.

Noise words ensure that your search lists will not be cluttered with unhelpful or forbidden terms and preserve the speed of retrieval of indexed records.

Common noise words:

- about
- and
- of
- that
- the

You can add as many of your own words as you want to the noise word list.



About OpenText

OpenText enables the digital world, creating a better way for organizations to work with information, on-premises or in the cloud. For more information about OpenText (NASDAQ/TSX: OTEX), visit opentext.com.

Connect with us:

OpenText CEO Mark Barrenechea's blog

Twitter | LinkedIn

Feedback

If you have feedback or queries about this document, connect with Content Manager Support.