

dtSearch Desktop/Network  
Indexing and Search techniques

## **T205 - SEARCH FOR TRANSLATORS**

dtSearch Desktop/Network is a powerful search tool used by professionals for a wide variety of tasks, this short tutorial aims to show you how to index and search Microsoft TBX and CSV glossaries. Although aimed at translators, the techniques will be found useful by other professionals that need to index very large text files.

### **Course Requisites**

dtSearch Desktop/Network 7.68 or later  
UltraEdit text editor.  
Internet access

## T205 - SEARCH FOR TRANSLATORS

This training course covers several advanced topics of interest to those that need to index and search Microsoft TBX and CSV Glossary files.

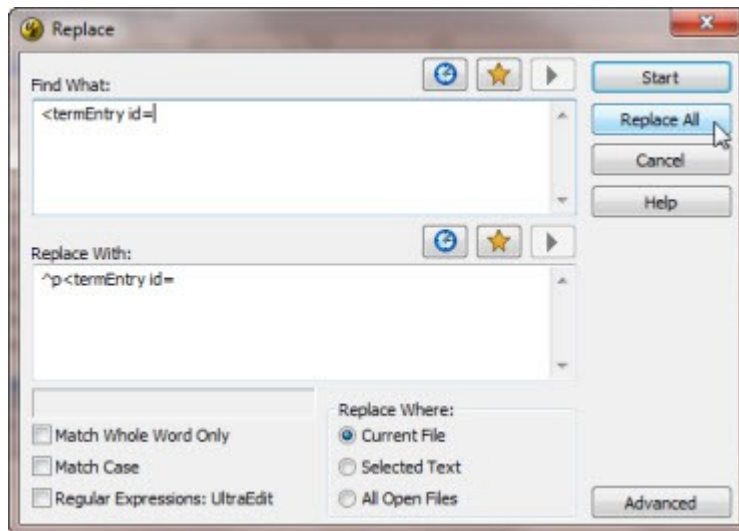
The first part of this tutorial covers indexing and searching a TBX (**TermBase eXchange**) file. TBX files are an XML file in an international standard format for the interchange of terminology data including detailed lexical information.

This TBX example is based on the YouTube Video <http://www.youtube.com/watch?v=YLfgH4iAo8s> by CATguruEN (Dominique Pivard)

If you do not have a suitable Microsoft TBX file on your PC, you will need to download a file from here: <http://www.microsoft.com/Language/en-US/Terminology.aspx> select French then click on the Download button. Edit the file name by adding `_FR` to distinguish it from any other files you may download later and Save the resulting `MicrosoftTermCollection_FR.tbx` file in your Downloads folder.

The file needs to be prepared so that it is easy to segment later in dtSearch. If you do not have a copy of UltraEdit you can download a 30-day trial version or purchase a copy of the latest version from: [http://www.ultraedit.com/downloads/ultraedit\\_download.html](http://www.ultraedit.com/downloads/ultraedit_download.html)

Run UltraEdit and open the Microsoft TBX file, locate the text `<termEntry id=` and select it with your mouse, from the Search menu select Replace.

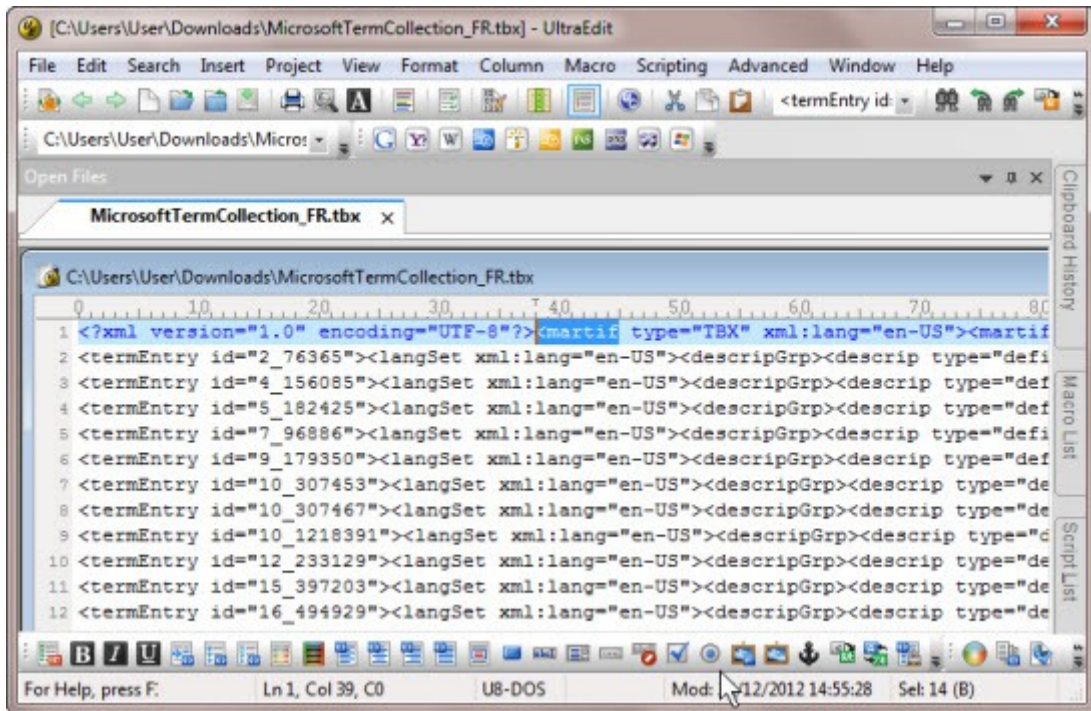


The Replace dialog will open with the selected text in the "Find What" text box. Copy the text into the "Replace with" text box and edit it to add `^p` at the start of the line as shown above.

Click on **Replace All**

This will format the text so that each line starts with the text `<termEntry id=` as shown below.

## T205 - SEARCH FOR TRANSLATORS

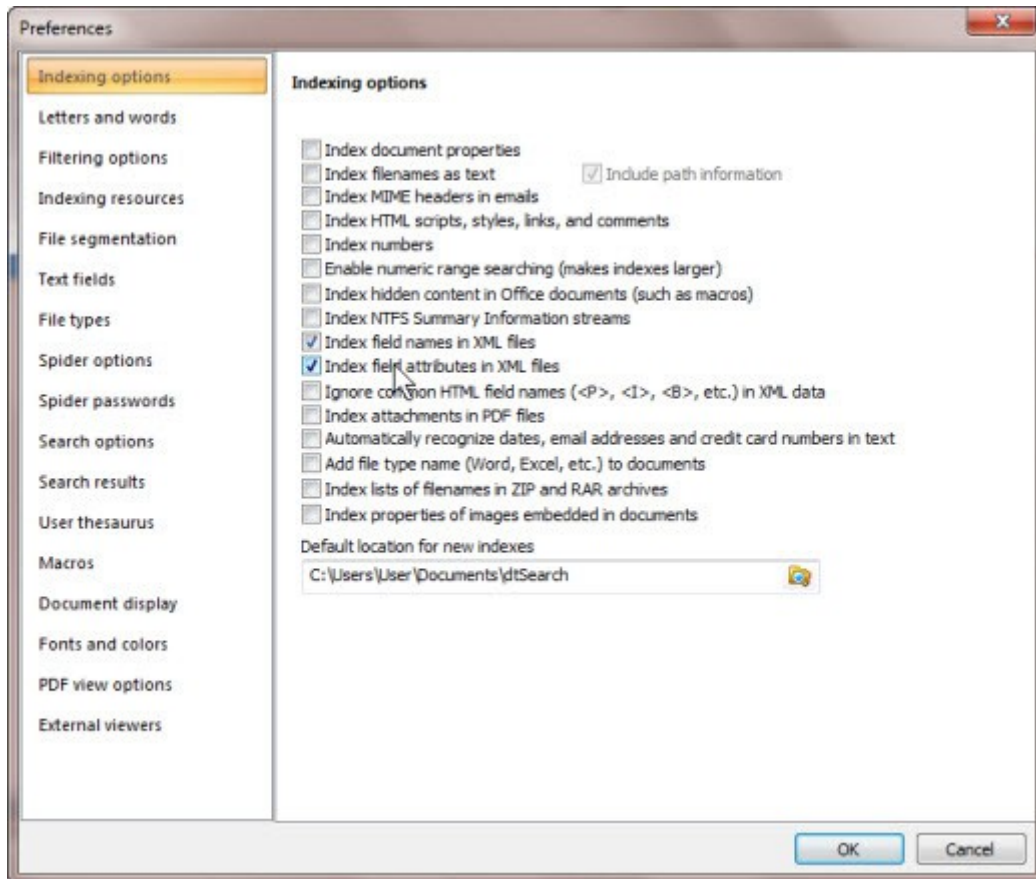


Now we are ready to set-up dtSearch Desktop.

**Initial set-up of dtSearch Desktop:**

From the **Options** menu, choose **Preferences > Indexing Options**

TBX files are XML files; check the "Index field names in XML files" and "Index Field attributes in XML files" check-boxes as shown below.

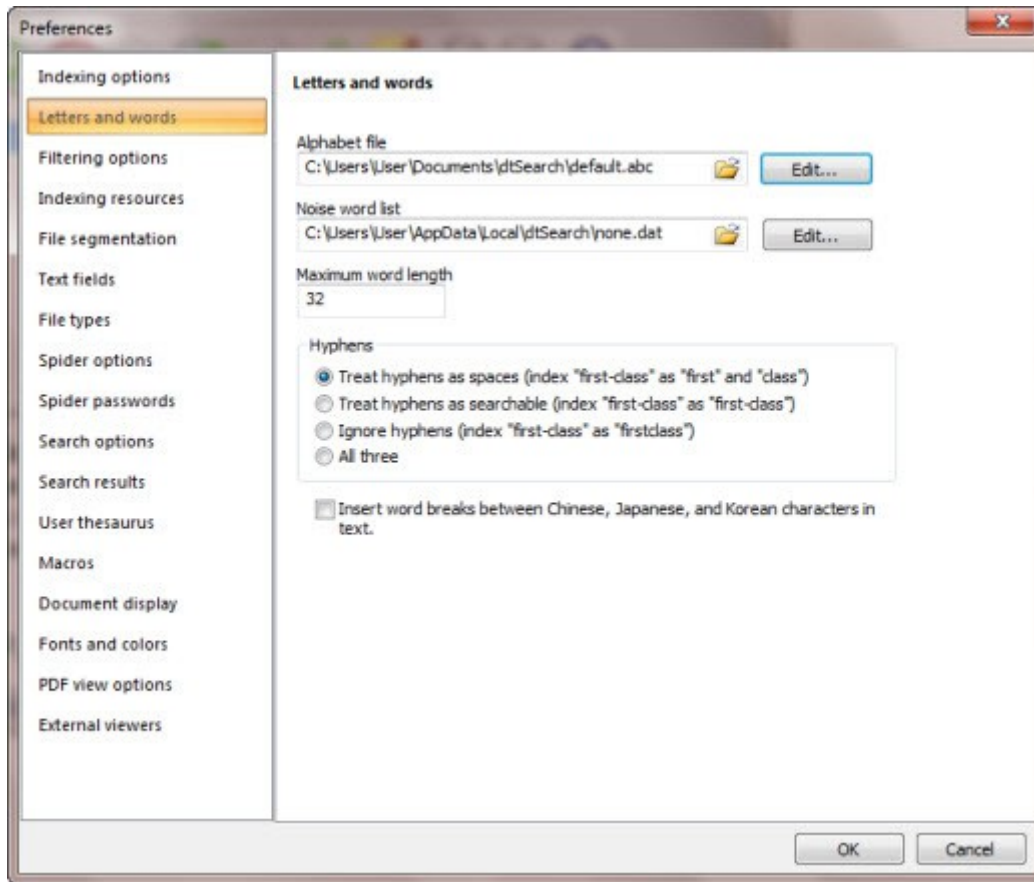


**TIP.** To use the keyboard instead of a mouse to navigate, use **Ctrl+Tab** or **Ctrl+Shift+Tab** to move down or back up in the left hand panel. Use the **Tab** key or **Shift+Tab** to move down or up in the right hand panel.

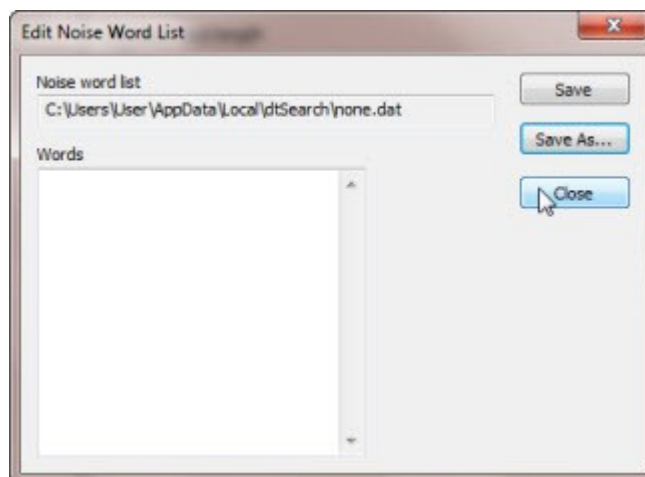
Next choose **Letters and words**

We need to make sure the default.abc Alphabet file has the factory default settings, click on the Alphabet file Edit button. Make sure that the \$ sign (ASCII 36 decimal) and all other characters from 33 to 47 are set to **Space**. If you make any changes click on the **Save button** before closing the dialog.

## T205 - SEARCH FOR TRANSLATORS



Now click on the **Edit...** button alongside the **Noise word list** text-box. For this session we need an empty noise word list. Create one by deleting all the words in the list, then click on the **Save As...** button and save it with a file name of none.dat, now **Close** the dialog.



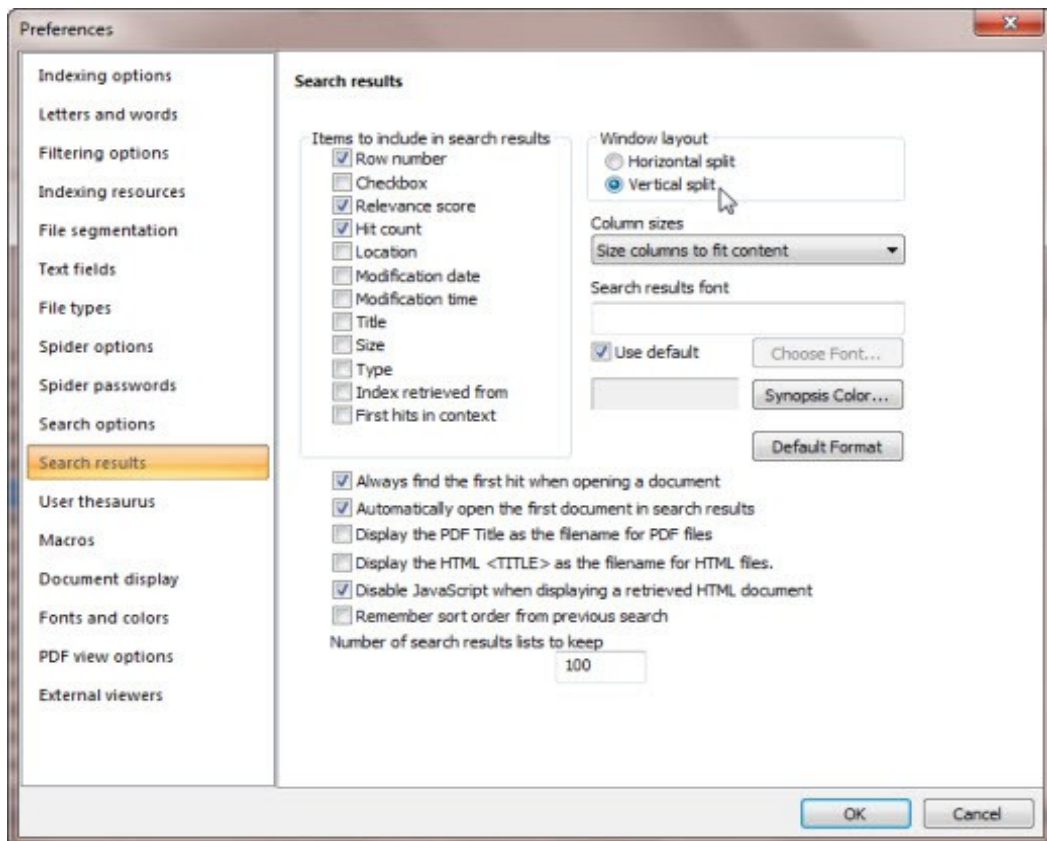
Now select **Search results** from the left hand pane.

## T205 - SEARCH FOR TRANSLATORS

From the **Search results** panel check the Row number, Relevance score and Hit count check-boxes as shown below.

Since we are only displaying results with a few columns from a single file a vertical split will be found more convenient.

Finally, check the **Always find the first hit when opening a document** and **Automatically open the first document in search results** checkboxes.

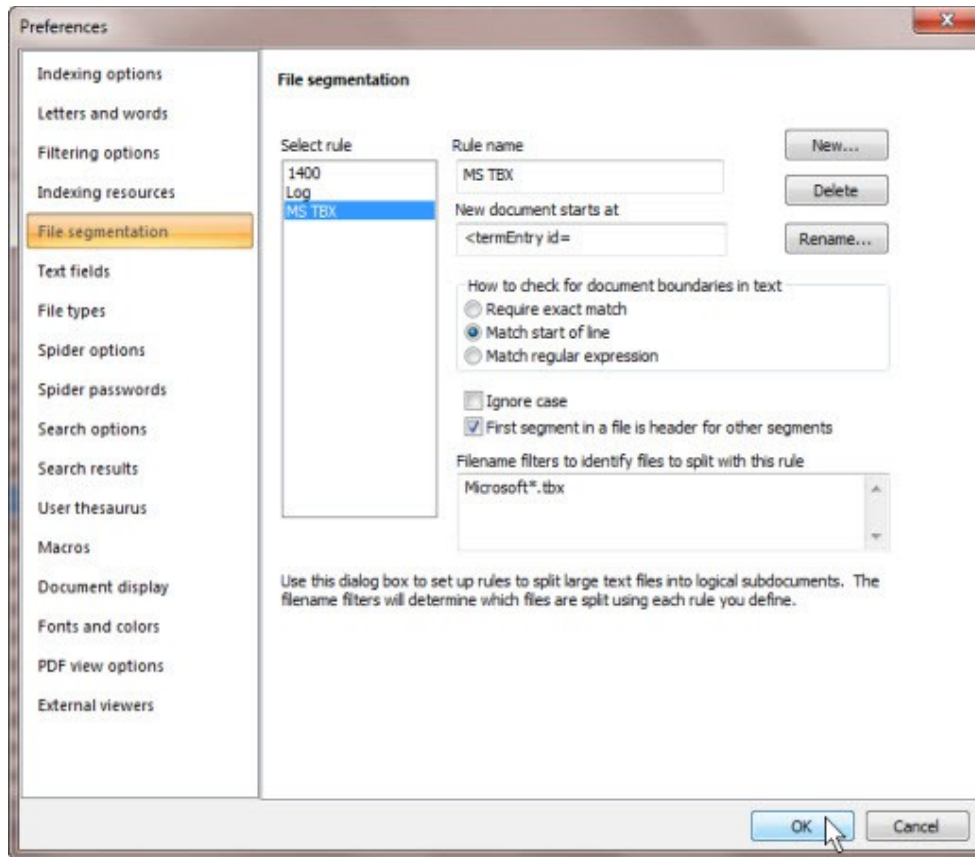


Now select **File segmentation**



## T205 - SEARCH FOR TRANSLATORS

In the File segmentation panel press the **New...** button and enter a rule name **MS TBX** as shown.



It is very important that the text entered in the "New document starts at" text box is an exact match to the text in the file, to avoid error it is best to paste the text in that was copied from the TBX file earlier.

Select the "Match start of line" and check the "First segment in a file is header for other segments" .

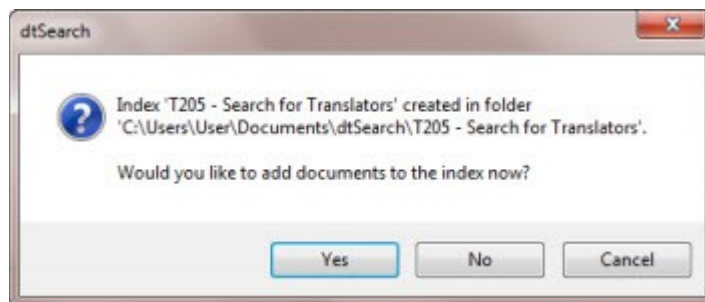
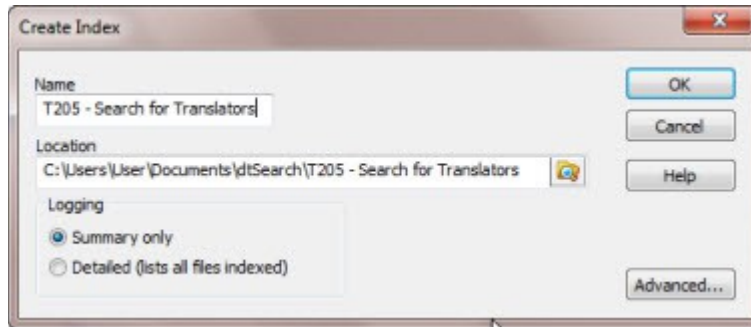
Enter **Microsoft\*.tbx** in the File name filters to identify to split with the rule.

Finally click the OK button.

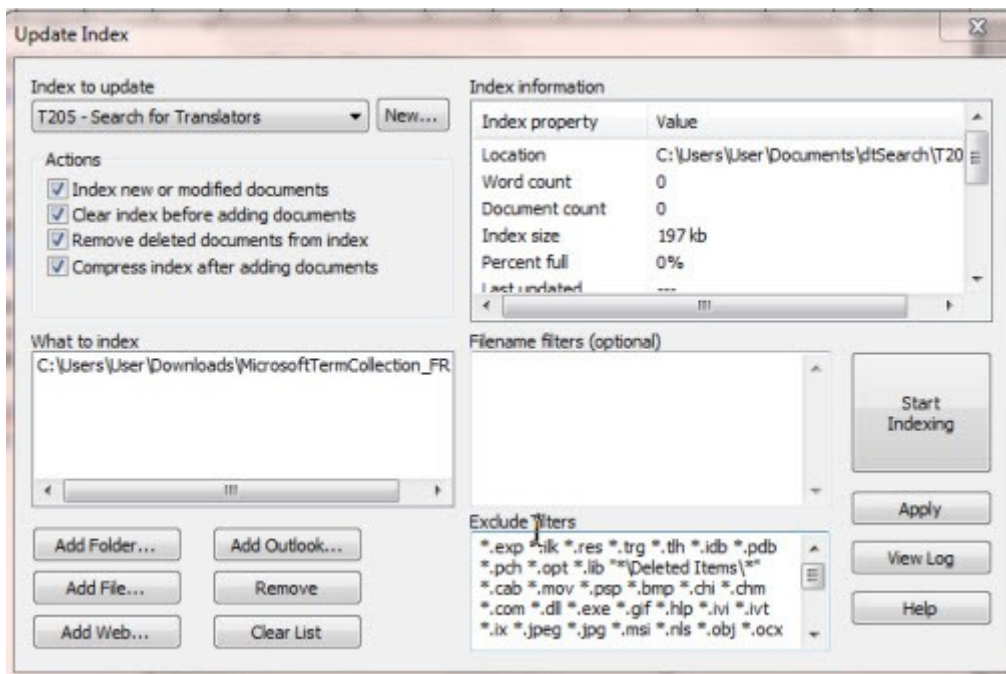
## T205 - SEARCH FOR TRANSLATORS

Now we are ready to create a basic index. From the **Index** menu select **Create index...**

Enter a name for the index as shown below:



In the **Update Index** dialog that appears, press the **Add File..** button, now browse to the Microsoft TBX file that has been reformatted.



The **Update Index** dialog will re-appear, press the **Start Indexing** button, when the indexing is complete (under 30 seconds on a broadband Internet connection) click on the **Close** button.

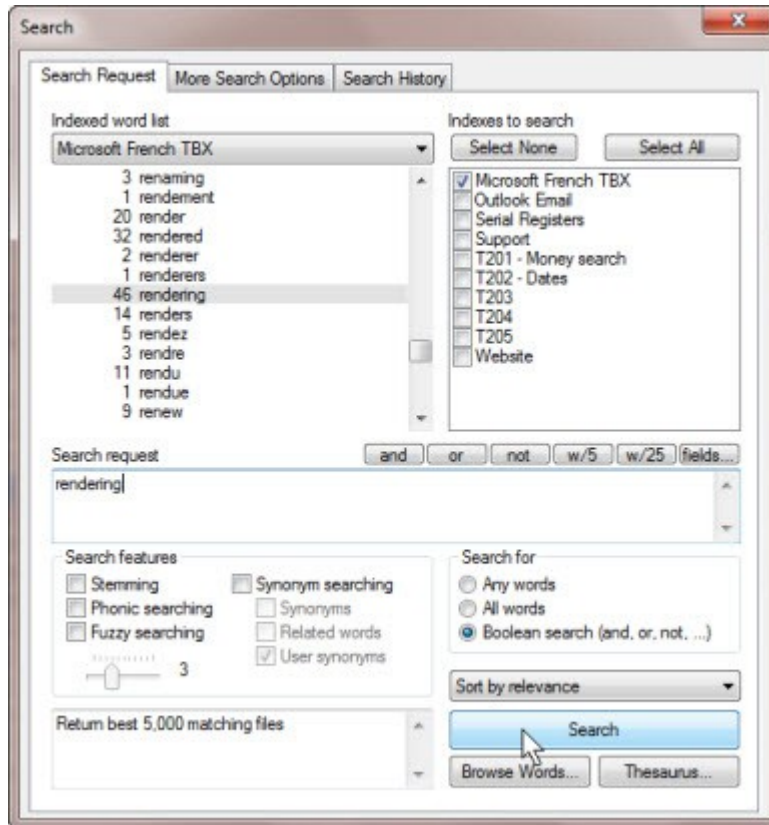
**We are now ready to start searching!**



## T205 - SEARCH FOR TRANSLATORS

In dtSearch Deskop click-on the Search icon or press Ctrl+S to open the Search dialog.

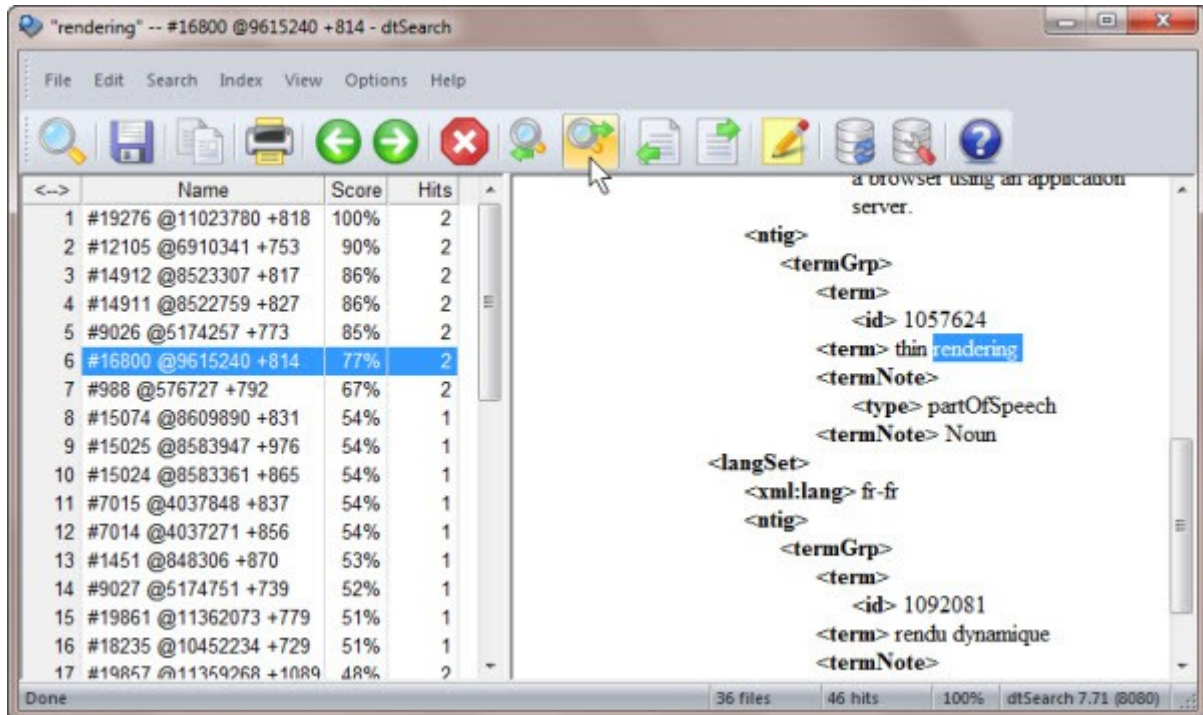
Press the **Select None** button to unselect any previously selected indexes then select the "Microsoft French TBX" index. Make sure no **Search features** check-boxes are selected and that **Boolean search** is selected.



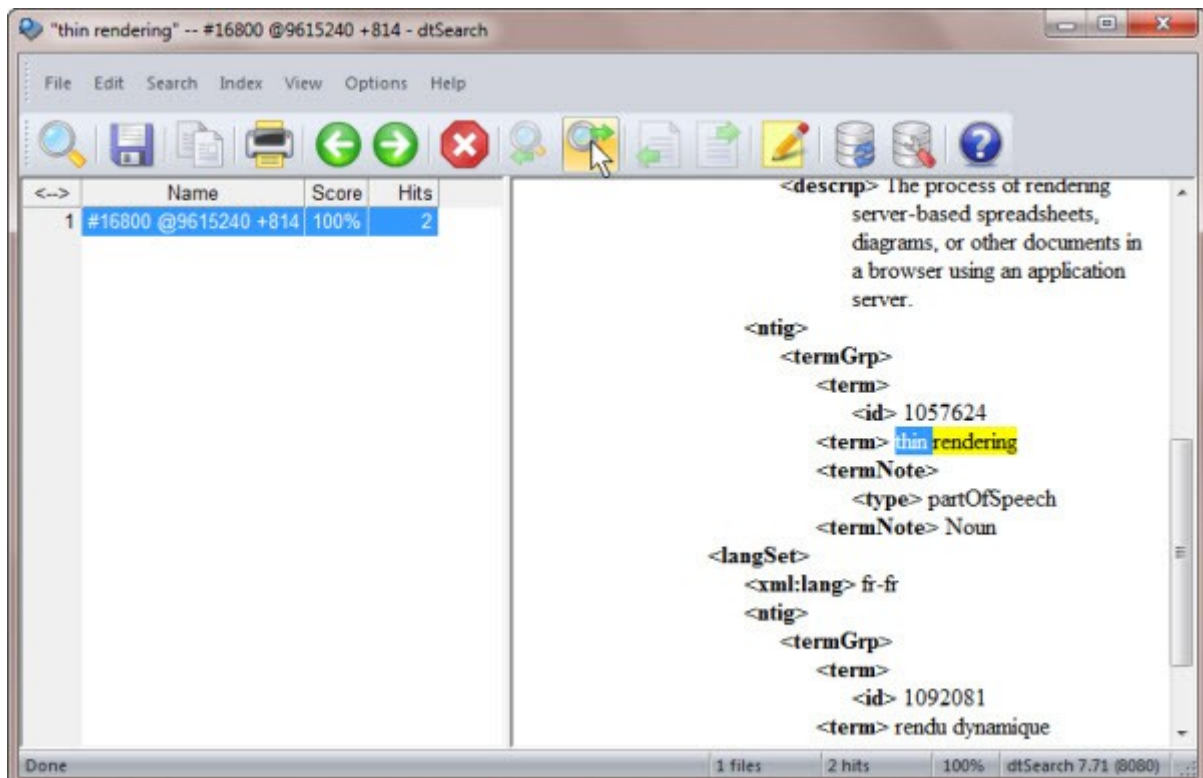
Enter your search request and click on the Search button.

## T205 - SEARCH FOR TRANSLATORS

Click on the Next Hit button until you find the result you are searching for, this example is taken from the YouTube video by CatguruEn and shows the term "thin rendering" the translation of the French term "rendu dynamique".



The dtSearch Boolean search allows phrases to be entered without the need for quotation marks, repeat the search using the search request `thin rendering` to find the single exact translation.



## **Part 2**

### **How to create and search**

### **CSV Glossaries**

#### **Prerequisites**

OpenOffice Calc or similar spreadsheet software.

## T205 - SEARCH FOR TRANSLATORS

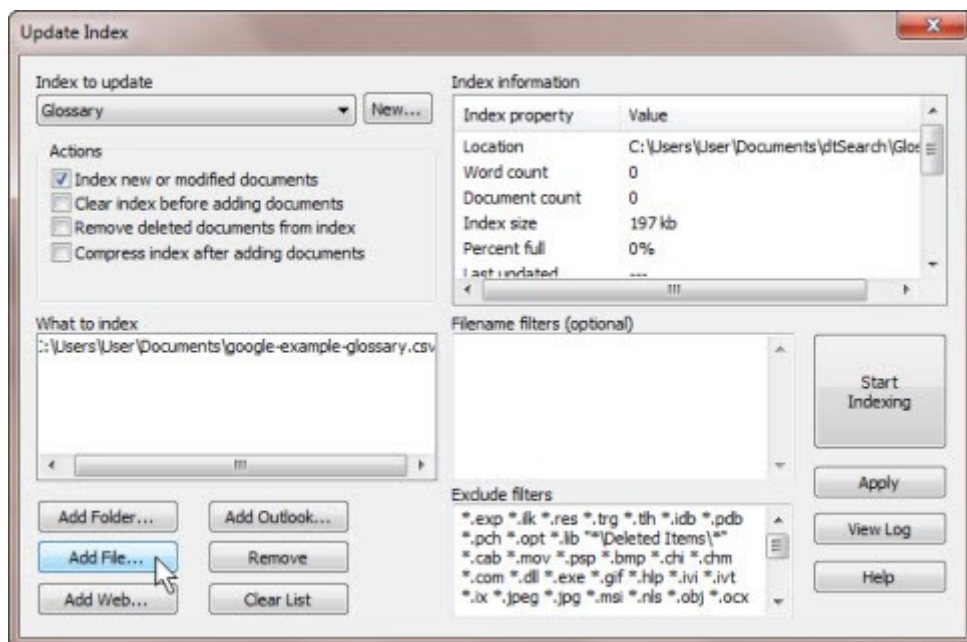
To create a CSV file in the Google Translator Toolkit's glossary format see the instructions at <http://support.google.com/translate/toolkit/answer/147854/?hl=en> use your own translated words or phrases or copy the example shown.

Alternatively you can download the file used in this tutorials example from <http://www.dtsearch.co.uk/training/T205/google-example-glossary.csv>

CSV files need no preparation prior to indexing, dtSearch treats each row as a separate document, provided that the CSV file has a .csv file name extension, a header row (list of field names in the first line), and each field is separated by a comma, semi-colon or tab character. See list of supported file types in the fields section <http://support.dtsearch.com/faq/dts0103.htm>

In dtSearch Desktop you will need to make sure that the alphabet file has the default settings and that the search result layout and index options have been set appropriately; if you have just carried out part 1 of this tutorial you need make no changes, otherwise set up dtSearch Desktop as shown on pages 3 to 5.

From the dtSearch Desktop Index menu select Create Index... , name it `Glossary` and when prompted click on the Yes button to add documents to the index. From the dialog that opens click on the **Add File...** button and browse to the CSV file that you have downloaded or created.



Since this is a new index just select the **"Index new or modified documents"** checkbox, then click on the **Start Indexing** button. Close the dialog when the indexing has finished (about 1 second).

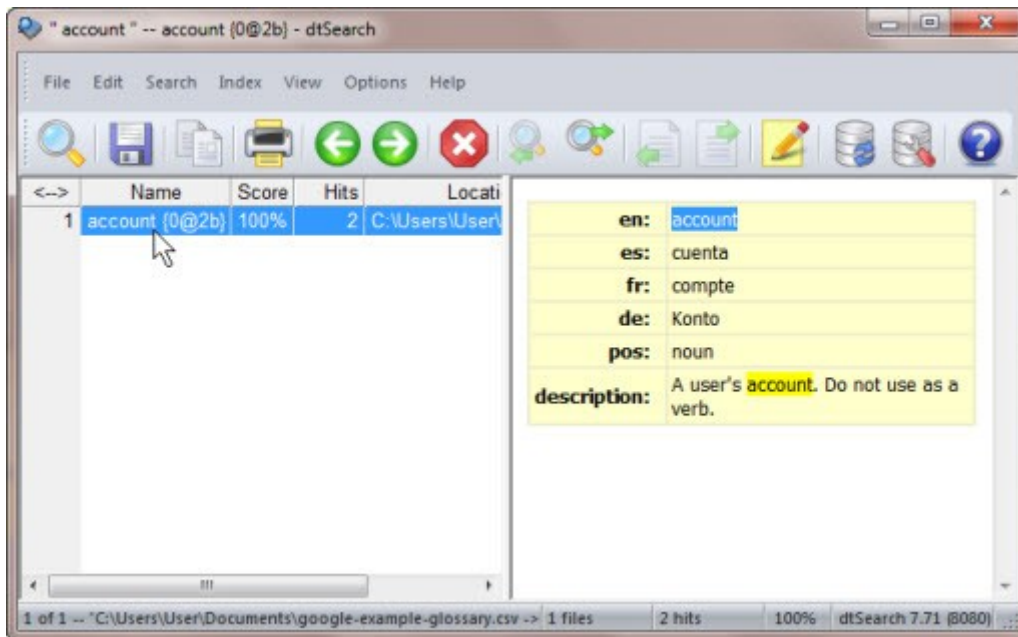
**You are now ready to search the glossary.**

## T205 - SEARCH FOR TRANSLATORS

Press Ctrl+S or click on the Search icon to open the Search dialog. Ensure Boolean is selected and that all other search features (fuzzy, stemming, phonic, synonyms) are not selected.

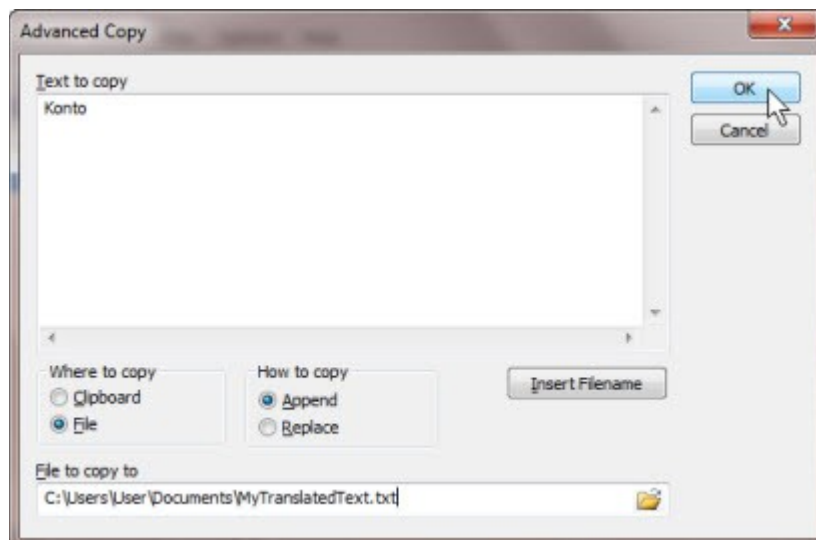
Now enter the search term `account` and press the search button.

The search result will show the search term highlighted together with the translated term(s).



### Tip:

You can select the translated text then **right-click > Copy** and then paste into your document in the usual way or you can select the translated text, then press **Ctrl+K** (or from the **Edit** menu select **Copy Advanced...**) the text will appear in the "text to copy" window, you can then append the text to your file.



Issue 0.2 -29 Dec 2012