

myCAT

Quote Detector User Manual

1. General Description

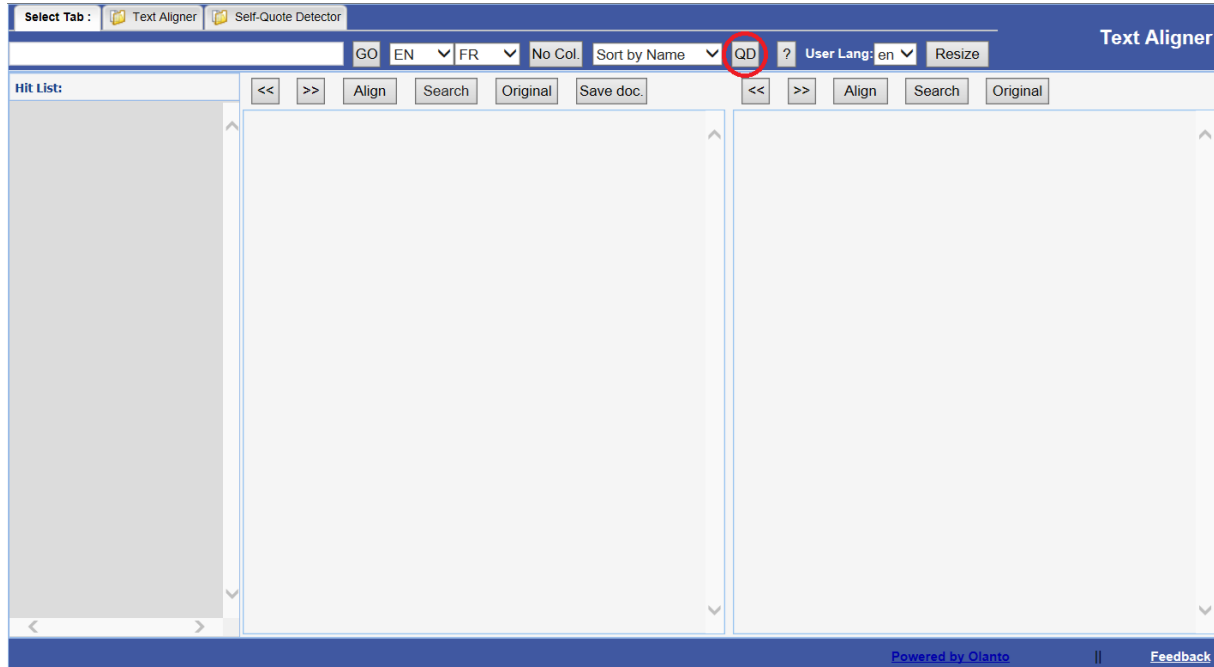
myCAT is a computer-assisted translation tool which includes several tools: a Text Aligner, a Quote Detector and a Self-Quote Detector. This document is the User Manual for the Quote Detector.

The Quote Detector is a tool which was designed to analyze a new document before it is translated, so as to identify all the parts (from a few words to entire paragraphs or pages) which already appear in a previous document, and for which a translation already exists.

2. Accessing the Quote Detector

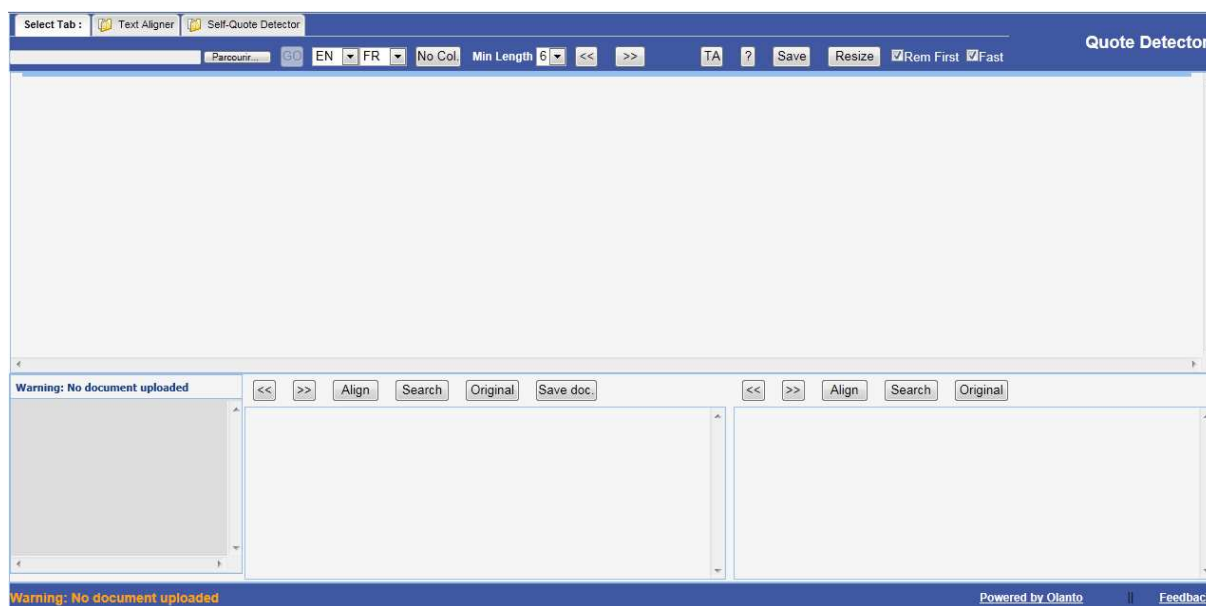
The Quote Detector is accessed from the interface of the Text Aligner by clicking on the “QD” button.

Supported Browsers: myCAT supports Internet Explorer 8 or later and Mozilla Firefox 3.6 or later.



3. Using the Quote Detector

The Quote Detector environment looks like this:



If you have been using the Text Aligner before using the Quote Detector, you can note that the language scroll-down menus and the “Filter OFF” collection button are still available; you can use them as described in the Text Aligner User Manual.

3.1. Uploading the Document to be Analyzed

To upload the document you want to analyze (that is, the document to be translated, for which you are looking for quotes), click on the “Browse” button in the top left corner. A standard Browse box will open.

In that Browse box, go to the location of the document to be analyzed and click on the “Open” button. This operation is completed when the analyzed text appears in the top frame with quotes displayed in yellow.

Notes:

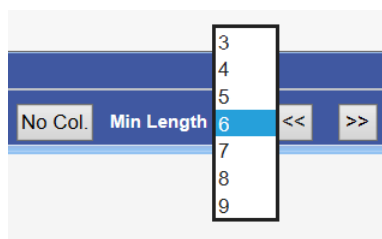
- The supported formats are doc, docx, pdf, html and txt.
- There is a **limit to the size of the documents which can be uploaded**: by default, this limit is set to **3’072 Kb**, *i.e.* 3 Mb. This is to avoid submitting very large documents which would unnecessarily overload the Quote Detector (remember that several users may upload a document at the same time). To remain under this limit, a simple solution is to delete most of the photographs, diagrams, etc. which may inflate the document’s size but will not be analyzed anyway. Another simple solution is to select the number of pages you are intending to translate today, and to copy/paste them into a provisional document which you can then submit to the Quote Detector.

3.2. Selecting the Source and Target Languages

Before launching the analysis, you can select which source and target languages you are working in by using the scroll-down language menu. You can also select one or several collections, if needed, from the “Filter OFF” button (see the User Manual of the Text Aligner for more details).

3.3. Setting Up the Minimum Length of Quotes and Flags

You can choose the minimum number of words you would like to see in the detected quotes by using the “Min. Length” scroll-down menu. This function avoids detecting very small quotes, which could be numerous but not very useful. The lowest length is 3 consecutive words; the highest is 9. Please note that this highest possible value only applies to the minimum number of words to appear in a quote; there is no maximum length for the quote. In fact the system always looks for the longest possible quote; thus if the whole document was already translated it should appear entirely in hyperlink.



Since version 3.7, you can choose to remove the first reference (the one that is most often cited) by checking "Rem First".

The checkbox “Fast” behaves as follows:

- when Checked: Selects the algorithm of previous versions (before 3.7), fast but with some errors
- when Unchecked : the algorithm checks references and eliminates errors (slower but more accurate)



3.4. Launching the Quote Detection

When all the parameters are set up you can click on the “GO” button to launch the analysis. The message “Uploading File...” appears in the bottom bar. The task is completed when the analyzed text is displayed in the top frame with quotations highlighted in yellow:

A screenshot of the 'Quote Detector' software interface. The top bar shows the file name 'C:\MYCAT\Test_doc.docx' and various settings like 'EN', 'FR', 'No Col.', 'Min Length 6', and buttons for 'GO', '1/3', 'TA', '?', 'Save', and 'Resize'. The main window is titled 'STATISTICS' and displays the analyzed text with several quotes highlighted in yellow. Below the text, there is a 'TOP' button and statistics for the document: 'Total number of words: 130', 'Number of references found: 3', and 'Number of words found in references: 106 (81.5%)'. At the bottom, there is a 'Hit List' showing two hits, each with a preview of the text and a button to view the full text. The bottom status bar shows 'Last Hit / 1', 'Powered by Otanto', and a 'Feedback' link.

In the example above, to test the Quote Detector, we introduced a completely new sentence among a set of referenced sentences. The application successfully stopped its quotes before the unknown sentence and started again after it.

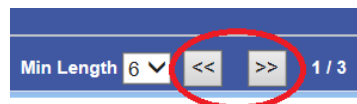
The quotes are numbered as [R1][E1], [R2][E2], etc., “R” being the beginning of reference tag and “E” the end tag. You can click on any quote to display the list of documents which contain it. (Note: the list of documents containing the first quote of the analyzed document is automatically displayed after the document analysis is completed.)

As with the Text Aligner, clicking on a reference file name in the Hit List displays the content of that document in the source and target versions, and aligns both documents on the quoted part.

You can then copy the corresponding part of the target text and paste it into your current translation work.

3.5. Using the Quote Navigation Button

The navigation buttons >> and << located at the top of the screen allow for an easy transition between the various quotes. The number of the quote currently displayed and the total number of quotes found for the document are also displayed next to the navigation buttons:



3.6. Using the Lower Frame Buttons

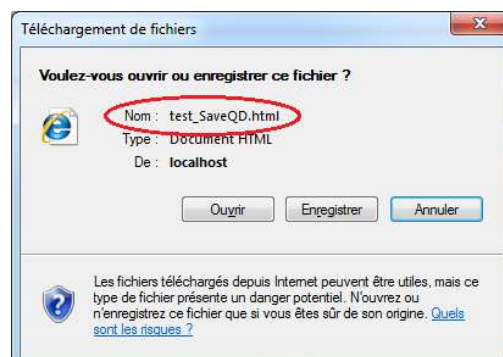
When the Quote Detector displays a reference document, several buttons are made available:



Those buttons are identical to those of the Text Aligner and are described in the Text Aligner User Manual.

3.7. Using the “Save” Button in the Top Bar

After a document was analyzed it is possible to save the resulting file with all its references (in HTML format) so as to revert to it later without having to perform the analysis again. To do so, after the analysis is completed, simply click on the “Save” button located at the top right of the screen. The browser will propose, in a pop-up window, to open or save the file. You can choose to save the HTML file to any location:



At a later stage, if you want to re-use the analyzed file, simply open it in the Quote Detector environment by using the “Browse” button. The HTML file will be displayed along with its references (the GO button is de-activated since the analysis is already performed). You can click on any reference and see the reference files in the bottom frames.

3.8. Understanding the Statistics in the File Created with the “Save” Button

As described in the previous section, the “Save” button creates a copy of the analyzed text in which it highlights all the quoted parts and for which it saves all the links to the reference document.

However, there is more information in this file: The Quote Detector builds, at the end of the analyzed text, a statistics section which includes the following elements:

- **General statistical information:**

Statistics of document C:\MYCAT\test\test.doc
Total number of words: 155
Number of references found: 6
Number of words found in references: 123 (79,4%)
Minimum of consecutive words: 4
Collection: All collections
Date: Fri Nov 23 12:14:33 CET 2012

- **Statistics on the reference documents**, along with their frequency of use and the reference number of the related quotations:

Reference Document	%	References
C:/MYCAT/corpus/txt/EN/small-collection\WIPO\PLT\plt_ce_v_5.doc.txt	22,6 %	4, 5,
C:/MYCAT/corpus/txt/EN/small-collection\WTO\WTO_Agreement.doc.txt	20,6 %	6,
C:/MYCAT/corpus/txt/EN/small-collection\WTO\Final_Act.doc.txt	20,6 %	6,
C:/MYCAT/corpus/txt/EN/small-collection\UNO\A_RES_53_144.pdf.txt	15,5 %	1,
C:/MYCAT/corpus/txt/EN/small-collection\UNO\HR_Declaration.docx.txt	11,0 %	3,
C:/MYCAT/corpus/txt/EN/small-collection\UNO\A_RES_63_25.pdf.txt	9,7 %	2,

- **A list of all the quotes found**, along with their respective number in the analyzed text and the path and name of their reference document(s):

Reference Number	Referenced Text	Reference File
1	Declaration on the Right and Responsibility of Individuals, Groups and Organs of Society to Promote and Protect Universally Recognized Human Rights and Fundamental Freedoms	C:/MYCAT/corpus/txt/EN/small-collection/UNO/A_RES_53_144.pdf.txt
2	Reaffirms its decision to commemorate the sixtieth anniversary of the Universal Declaration of Human Rights	C:/MYCAT/corpus/txt/EN/small-collection/UNO/A_RES_63_25.pdf.txt
3	Everyone has the right to form and to join trade unions for the protection of his interests	C:/MYCAT/corpus/txt/EN/small-collection/UNO/HR_Declaration.docx.txt
4	view of the Delegation, the Committee had made very good progress in developing a draft text for the Patent Law Treaty	C:/MYCAT/corpus/txt/EN/small-collection/WIPO/PLT/plt_ce_v_5.doc.txt
5	felt that in most areas one was getting close to an acceptable final text	C:/MYCAT/corpus/txt/EN/small-collection/WIPO/PLT/plt_ce_v_5.doc.txt
6	consensus is not reached at a meeting of the Ministerial Conference within the established period, the Ministerial Conference shall decide by a two-thirds majority of the Members whether to submit the proposed	C:/MYCAT/corpus/txt/EN/small-collection/WTO/Final_Act.doc.txt C:/MYCAT/corpus/txt/EN/small-collection/WTO/WTO_Agreement.doc.txt

3.9. Using the “Resize” Button

4. It is not always easy to adapt myCAT’s interface to the size of a computer’s screen. In order to facilitate and accelerate this process, a “Resize” button is available on the top bar. This button

instantly performs an automatic “Fit to Window” operation with regards to the current size of your browser. If your browser is extended to the whole screen, the myCAT interface will use the complete available screen. However you can also decide to reduce the size of your browser (especially if you have a large screen); clicking again on the “Resize” button will adapt myCAT’s interface to the new dimension of the browser.

4.1. Using the “Help” Button

This User Manual can be accessed at anytime from the Quote Detector interface by clicking on the Help or “?” button at the top of the screen.

4.2. Using the “Text Aligner” Button

A “Text Aligner” button is available from the Quote Detector interface in order to return to the Text Aligner application.

4.3. Links in the Bottom Bar

Two links are provided at the bottom right corner of the screen, as illustrated :



- **Powered by Olanto** : This link points to the web site of the Olanto Foundation, which is the publisher of myCAT (see <http://olanto.org> ; all the software published by Olanto is distributed under the AGPL open source license).
- **Feedback** : This links allows you to send an email in order to provide your feedback about myCAT. The default email address is info@olanto.org; it should be changed for the email of your Translation Support Unit.