



WWW Search Interfaces for Translators

The Internet—A powerful terminology research tool for translators

Tanya Harvey Ciampi

Do you ever find yourself spending far too much time on the Internet searching for a clue as to the meaning of a particular term or possibly even a translation? If so, here is a tool that can help put an end to that.

Anyone with any web experience will agree that the Internet is a powerful research tool for translators; many swear that they could not imagine working without it today (myself included!). Having worked as an English translator for a number of years, I was pleasantly surprised in 1997 to discover that there was a new way of finding terminology that did not involve setting out on a time-consuming trip to the library, making a series of more or less productive telephone calls or disturbing colleagues of mine who were up to their ears in their own work: This new research method involved the use of Internet “search engines”.

However, after the novelty wore off, time and again I found myself spending more time than necessary researching terminology on the Internet. It gradually dawned on me that I had to develop some search techniques capable of narrowing my search and speeding up the search process. As my knowledge of the various search engines grew, I put together a series of **search strings** that I could use every time I needed to find a definition or a translation of a term (that is, glossaries or parallel texts). Soon I realized that translator colleagues around the world with whom I had shared these techniques were finding them very useful, too, though cumbersome since they had to carefully type in the same long search string every time, bearing in mind that different search engines have different search syntaxes.

Eventually I came up with a way of automating the search process so that all you needed to do was enter your search term, and the rest was done automatically. Thus the **WWW Search Interfaces for Translators** were born. These tools are available online in various languages and are an ongoing project.

WWW Search Interfaces for Translators—www.multilingual.ch

The WWW Search Interfaces for Translators are designed to help you find **glossaries**, **parallel texts** and **lists of abbreviations**—that is to say definitions and translations of terms, and the long forms of abbreviations—in various languages. The interfaces are currently available in the Google version (best technique for finding glossaries) and the Altavista version (best technique for finding parallel texts), but others may be added in the future depending on the effectiveness of their search features.

Narrowing your search

The WWW Search Interfaces for Translators automate the search process by adding a string of **search criteria** to narrow your search, yielding fewer and more relevant results than a standard search, thereby reducing your search time.

Your search results are filtered using **fields** (in particular the **title**, **URL** and **link**) and **Boolean operators** “AND” and “OR”.

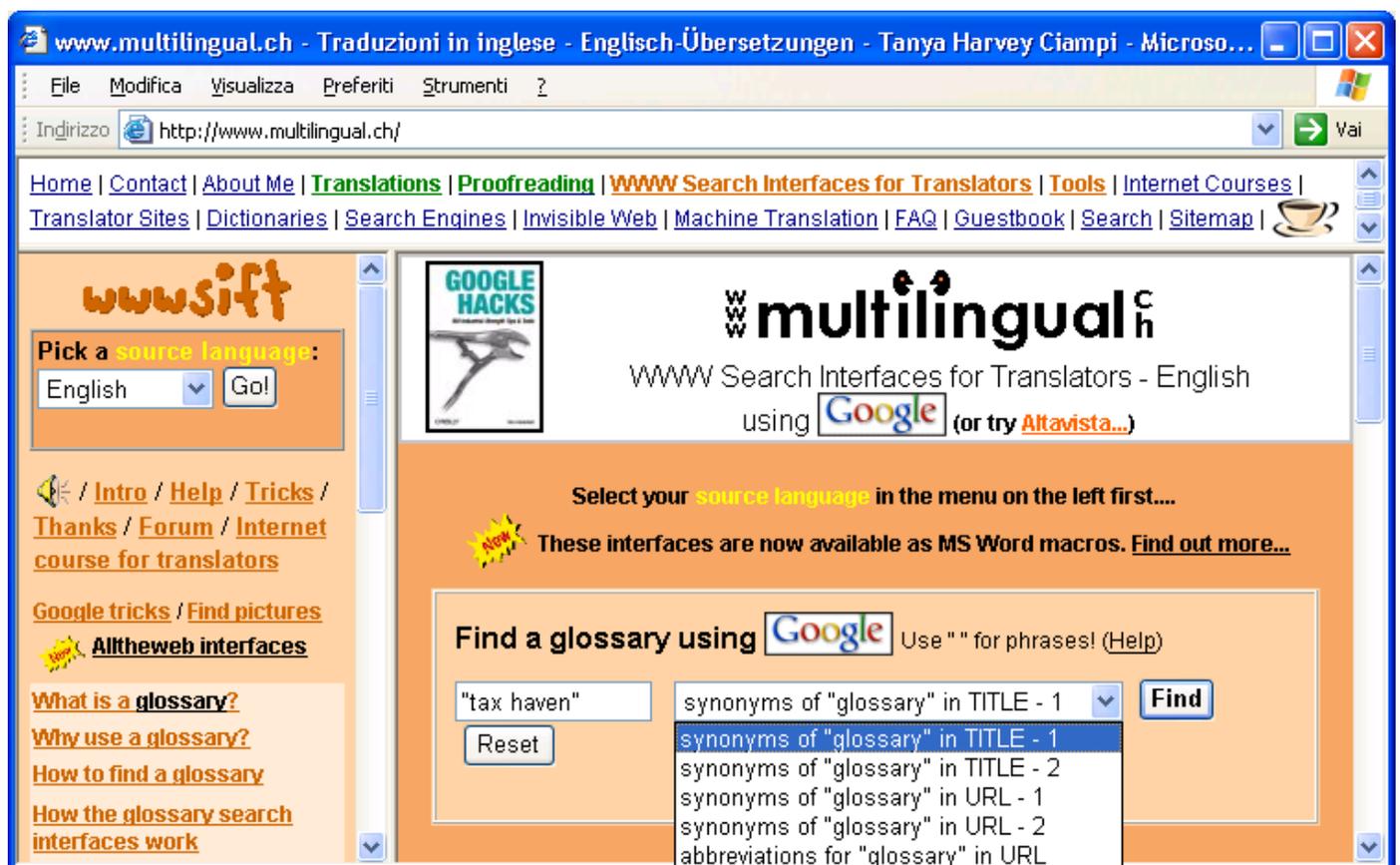
The results speak for themselves:

Example:

A standard search for the term “tax haven” yields **55,300 results**, few of which are relevant; With the search interfaces, we find **380 results** (*0.7% of the results above*), mostly relevant.

With the **Google glossary search interfaces**, the procedure is as follows:

Example of a glossary search:



The screenshot shows a web browser window with the URL <http://www.multilingual.ch/>. The page features a navigation menu at the top with links like Home, Contact, About Me, Translations, Proofreading, WWW Search Interfaces for Translators, Tools, Internet Courses, Translator Sites, Dictionaries, Search Engines, Invisible Web, Machine Translation, FAQ, Guestbook, Search, and Sitemap. The main content area has a logo for 'multilingual.ch' and the text 'WWW Search Interfaces for Translators - English using Google (or try Altavista...)'. Below this, there is a section titled 'Find a glossary using Google' with a search box containing 'tax haven' and a dropdown menu showing search results for 'glossary' in various fields like TITLE, URL, and ABBREVIATIONS. A 'Find' button is visible next to the search box. On the left side, there is a sidebar with a 'Pick a source language' dropdown set to 'English' and a 'Go!' button. Below that, there are several links and sections including 'Intro / Help / Tricks / Thanks / Forum / Internet course for translators', 'Google tricks / Find pictures', 'Alltheweb interfaces', and 'What is a glossary? Why use a glossary? How to find a glossary How the glossary search interfaces work'.

1. Enter your search term in the white box, using quotation marks (“ ”) for phrases;
2. Select a search technique (1 being the most effective) and hit the Find button;
3. Scan the search results for promising glossaries (see page title and URL).

Result: The following search strings are generated depending on the technique used
(**Note:** Google has a 10-word limit (“intitle” and “inurl” are not considered words) so your search term may consist of up to 7 words):

1. “tax haven” **intitle:glossary OR intitle:dictionary OR intitle:lexicon**
2. “tax haven” **inurl:glossary OR inurl:dictionary OR inurl:lexicon**
3. “tax haven” **inurl:dict OR inurl:gloss OR inurl:glos OR inurl:dic**

...

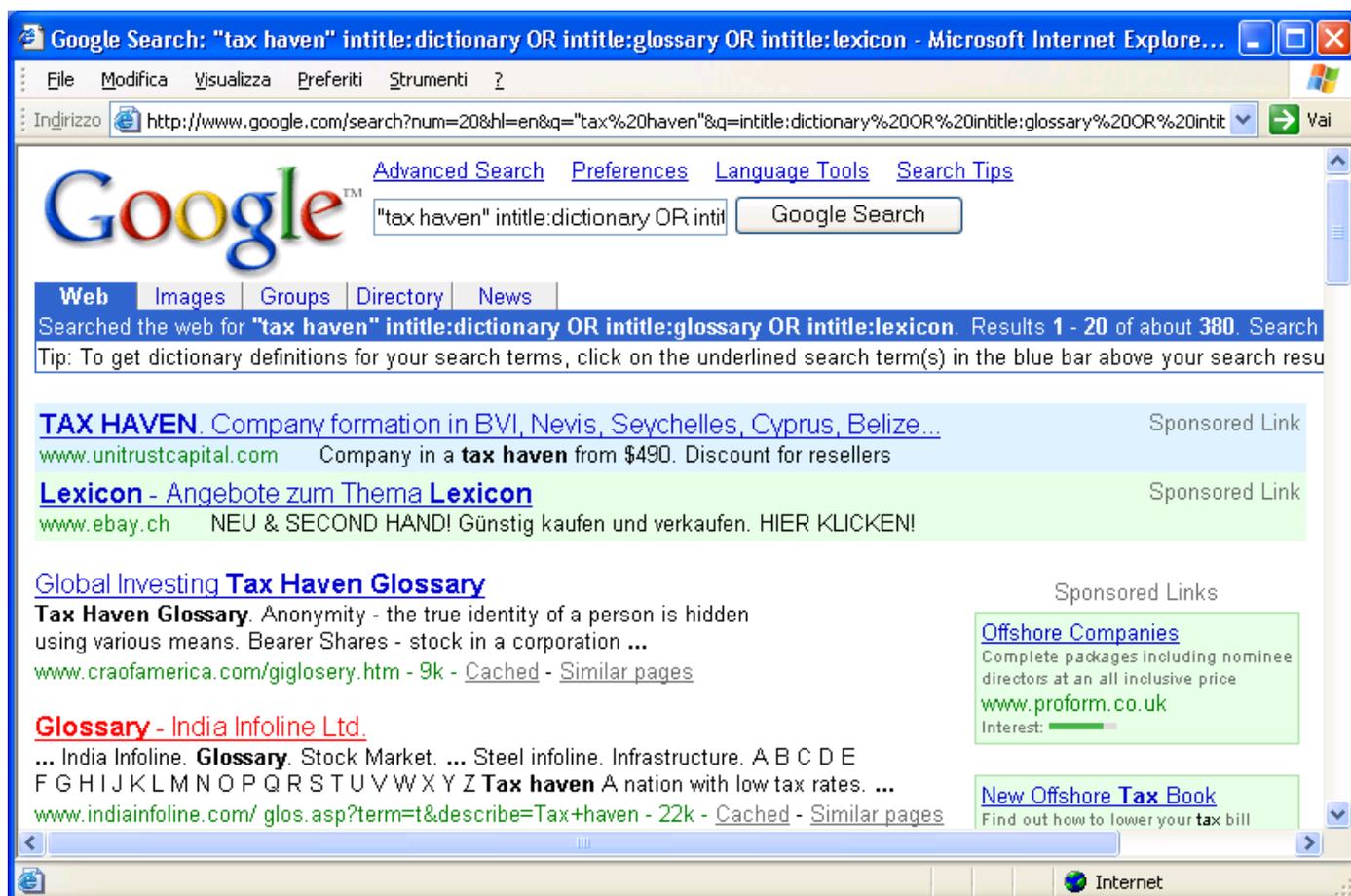
We might make the following comparison:

Title on the *cover* of a book = **Title** of a web page

Text on the *spine* of a book = **URL** (internet address) of a web page

Like most paper dictionaries, most on-line dictionaries and glossaries contain the word “dictionary”, “glossary” etc. on their “cover” (title) and/or “spine” (URL). The search strings generated exploit that fact as they search for your term, restricting the search to pages containing the word “dictionary”, “glossary” etc. in their title or URL.

Google search results: Scan the search results for promising pages (see title and URL)

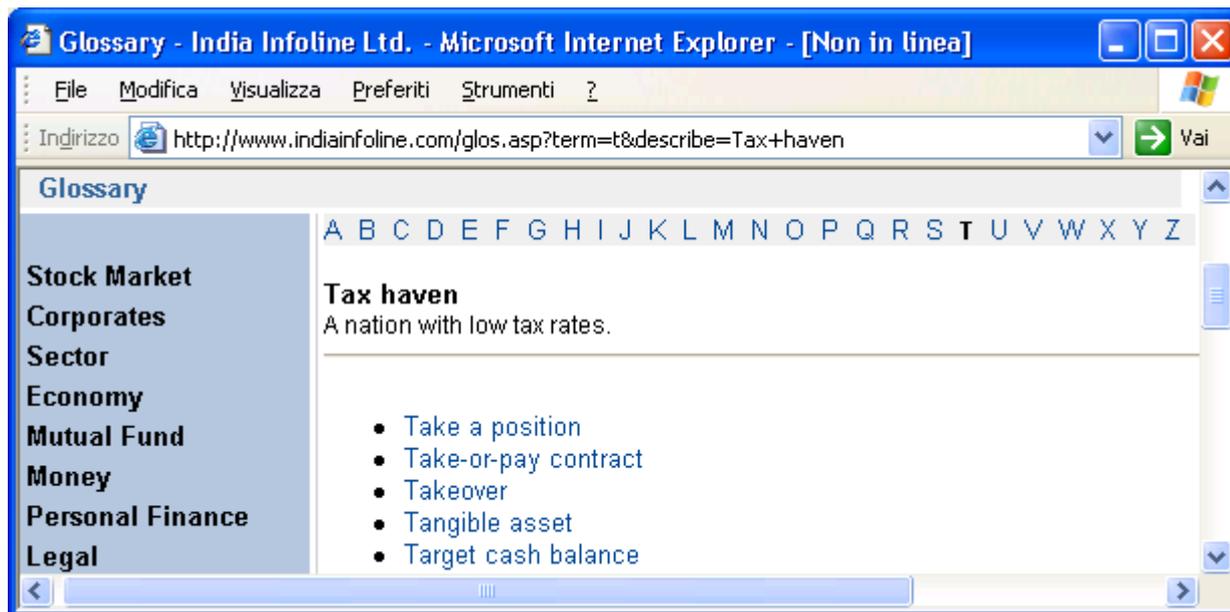


The screenshot shows a Microsoft Internet Explorer browser window displaying Google search results for the query "tax haven" intitle:dictionary OR intitle:glossary OR intitle:lexicon. The search bar contains the query, and the results show several sponsored links and organic search results. The first sponsored link is for "TAX HAVEN. Company formation in BVI, Nevis, Seychelles, Cyprus, Belize..." from www.unitrustcapital.com. The second is for "Lexicon - Angebote zum Thema Lexicon" from www.ebay.ch. The organic results include "Global Investing Tax Haven Glossary" from www.craofamerica.com and "Glossary - India Infoline Ltd." from www.indiainfoline.com. There are also two more sponsored links on the right side: "Offshore Companies" from www.proform.co.uk and "New Offshore Tax Book" from an unspecified source.

Note the most significant parts of the web pages in the list of search results:

- **title** (the underlined text in blue, a hyperlink to the actual page);
- **URL** (small green text below the title and description; it usually begins with “www”).

Example of a glossary found:



Note the most significant parts of the web page:

- **title** (the white text in the blue bar at the top of the page): contains the word “Glossary”;
- **URL** (black text in the white address bar below the title): <http://www.indiaonline.com/glos.asp?term=t&describe=Tax+haven>... (notice “glos” in the URL).

Why search the Internet for glossaries?

Glossaries can be a valuable source of terminology: Many glossaries available on the Internet contain terms that cannot be found in any commercially available paper or CD-ROM dictionary, either because **publication via the Internet is that much faster** than the traditional dictionary printing process—and thus better capable of keeping abreast with technology and neologisms—or because the terminology is **company specific** and would not normally be printed for the general public.

WWW Search Interfaces for Translators vs. specific paper or electronic dictionaries

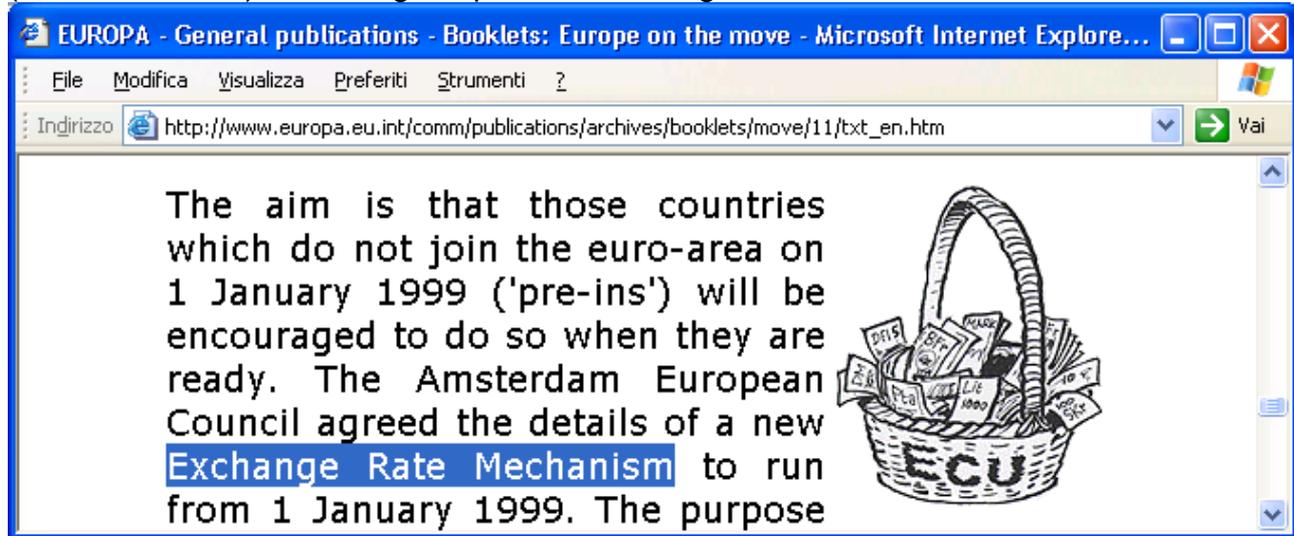
When you refer to specific paper or electronic dictionaries, you may find yourself having to search many dictionaries before finding your term—if you find it; with the search interfaces, the glossaries that you find *will* contain your term as *that* is the condition that is being set.

How reliable is information found on the Internet?

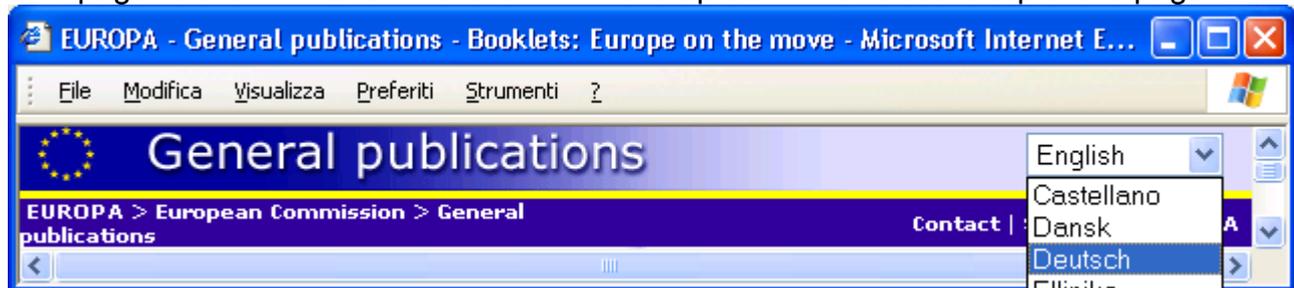
It goes without saying that it is ultimately up to you, the translator, to assess the quality and reliability of any information that you find on the Internet. You are advised to double-check the information that you find and consider the source.

What are parallel texts?

Parallel texts are corresponding texts in two or more languages. Below is a page in English (see “en” in URL) containing the phrase “Exchange Rate Mechanism”:



This page has “en” in its URL and a link in a drop-down menu at the top of the page...



... to a page with “de” in its URL, suggesting a corresponding page in German:



Parallel texts may be considered part of a large online multilingual corpus or translation memory database: they may be used for finding translations of **set phrases** (e.g. *solange der Vorrat reicht => fino ad esaurimento scorte, while stocks last etc.*) as well as **technical terms**.

Why use parallel texts?

Parallel texts can be a valuable source of terminology as they provide **translations of terms in context**, unlike most bilingual dictionaries.

Bilingual and multilingual dictionaries often provide many different translations for a specific term, leaving you, the translator, with the task of choosing the correct translation in your particular text, whereas parallel texts provide a specific translation in a specific context.

You might consider resorting to parallel texts when your **general** or **technical bilingual dictionary** is of little or no help.

How do the parallel text search interfaces work?

The same sort of filtering takes place here, too, but based on different criteria or search strings.

What are the distinguishing features of parallel texts?

- A language code in the URL as a directory name or file prefix or suffix (e.g. www.xxx.ch/deutsch/page.htm, www.xxx.ch/page_de.htm) indicating the language of the page in question;
- A hyperlink—in the form of text or a flag—featuring a language code (e.g. a link to pages www.xxx.ch/italienisch/page.htm or .../italiano/page.htm or the link “*Italian version*”).

The presence of one or both of these elements suggests that the page in question might exist in other languages, too, possibly your target language.

With the **Altavista parallel text search interfaces**, the procedure is as follows:

1. Choose a language, go to the Altavista interface page, select a technique and click Find;
2. In the Altavista window that opens, replace *keyword* with your source language term or phrase, e.g. *solange der Vorrat reicht*;
3. Hit the Search button;
4. Look through the search results for likely parallel texts (ignore pages such as www.xxx.de/page.htm, where “de” is not a *language code* suggesting the existence of the same page in other languages but simply a *domain name* indicating that the page is on a website registered in Germany (as opposed to www.xxx.de/de/page-de.htm).

Result: The following search string is generated:

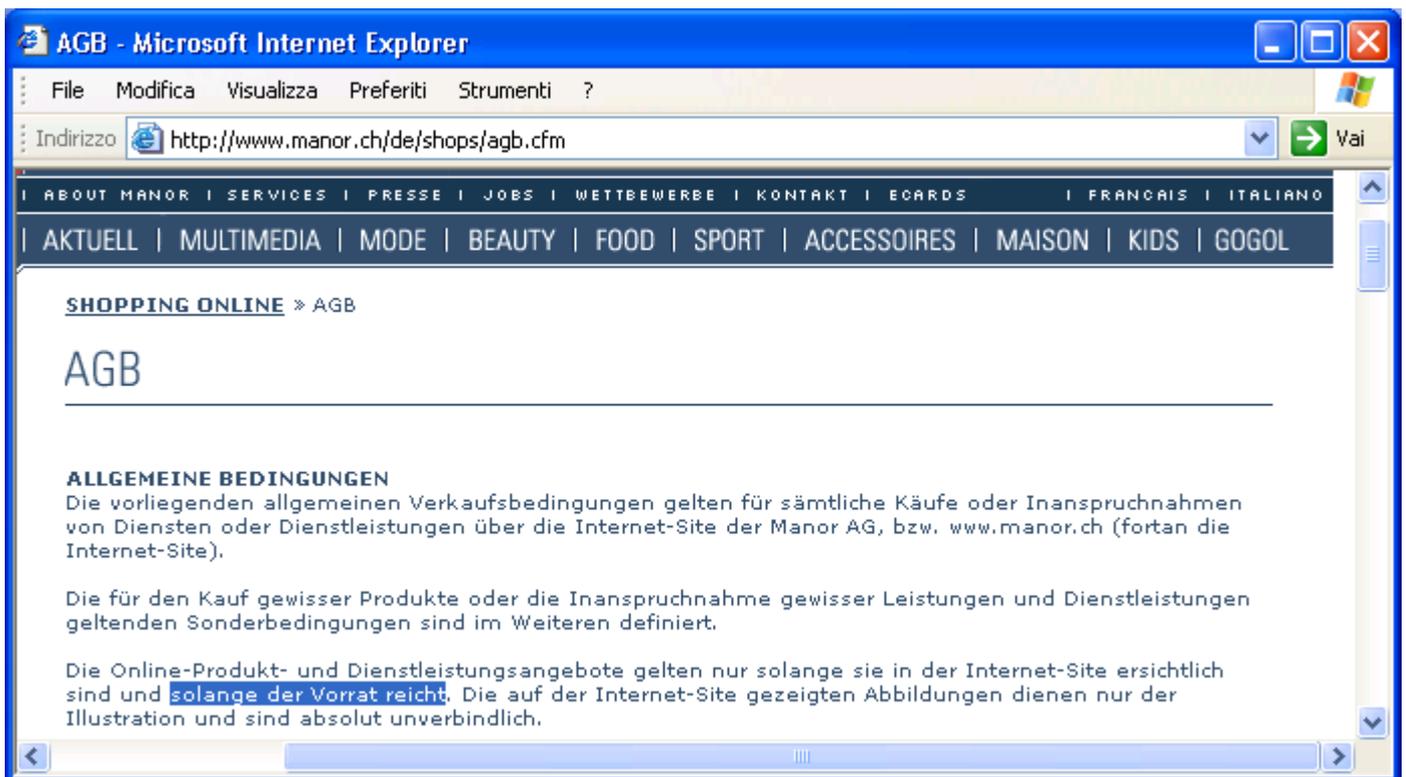
Example:

```
solange der Vorrat reicht AND (((url:d OR url:g) AND link:i) OR ((url:de OR url:ge) AND link:it) OR ((url:ger OR url:deu OR url:germ OR url:deut) AND (link:ita OR link:ital)) OR ((url:tedesco OR url:deutsch OR url:german OR url:allemand) AND (link:italiano OR link:italienisch OR link:italian OR link:italien)))
```

This looks very complicated, but all it means is “Find me a page that meets these criteria:

- 1) The page must contain my phrase (*solange der Vorrat reicht*);
- 2) It must have a language code in the URL indicating “German” as the language of the page;
- 3) It must have a link containing a language code suggesting a destination page in Italian.”

Example of pages found: The page below contains the phrase “solange der Vorrat reicht” and has “de” in its **URL** and a **link** (see top right corner) to a page containing “it” in its URL.



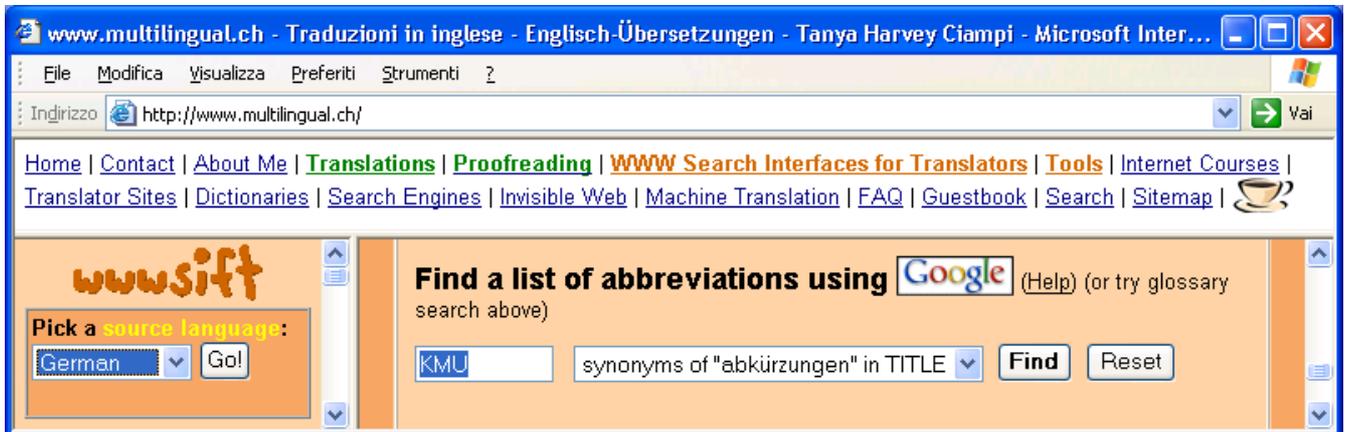
Corresponding page in Italian (see “it” in URL)



The **abbreviation search interface** works in much the same way as the glossary search interface but is specifically for finding the long forms of abbreviations.

Example: What does the German abbreviation KMU stand for?

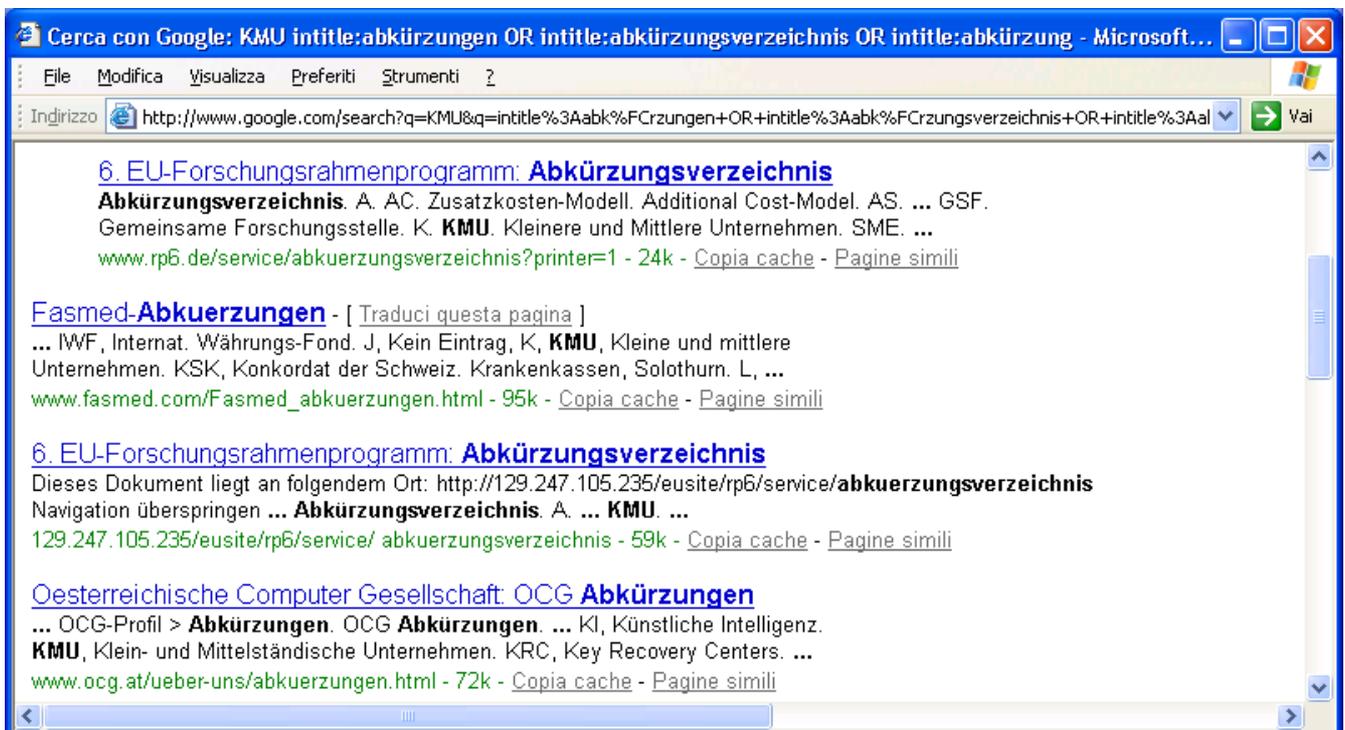
Select “German” in the menu on the left, enter your abbreviation and hit Find.



The following search string is generated:

KMU intitle:abkürzungen OR intitle:abkürzungsverzeichnis OR intitle:abkürzungsliste OR intitle:akronyme

Google search results:



A limited version of the WWW Search Interfaces for Translators is available in the form of **Microsoft Word macros**, allowing you to search for glossaries and parallel texts from within Word, making the search process even faster:



All you need to do is select a term and click either on a **glossary look-up macro** in the corresponding language to find a definition or on a **parallel text look-up macro** for the language combination in question to find a translation.

The **WWW Search Interfaces for Translators** enable you to spend less time surfing the Internet and more time getting on with your translation, increasing your productivity. The search interfaces in the form of **Microsoft Word macros** go one step further, reducing your search to a single click.

All feedback is welcome and will help improve these search interfaces in your particular language or language pair.

Courses and workshops on these search techniques are organized for translation teams in companies as well as at the Zurich University of Applied Sciences Winterthur (ZHW)—Department of Applied Linguistics and Cultural Studies, the former School for Translators and Interpreters (DOZ). For more information, visit my website.