

Internet search engines as auxiliary tools for translators

Svitlana M. Amelina¹, Rostyslav O. Tarasenko¹ and Vasyl D. Shynkaruk¹

¹National University of Life and Environmental Sciences of Ukraine, 15 Heroiv Oborony Str., Kyiv, 03041, Ukraine

Abstract. The paper focuses on using internet search engines as an auxiliary tool in the work of a translator. Emphasis is placed on the translator's skills in searching for information to support the translation process. Several search engines are discussed in terms of their use by translators to find additional and reference information on Internet resources. Aspects of the privacy policy of the systems in question are highlighted. The advantages of metasearch systems are outlined. A set of search options for translators has been defined – keyword searches, subject searches, and metadata searches. It is suggested to take into account the search engines of the countries for which the translation is being done.

Keywords: Internet search engines, information retrieval, translator

1. Introduction

The growing need for information in modern society in general and for each individual in particular is driving the emergence and development not only of new information and communication technologies, but also the complex transformation and expansion of the range of media content, as well as measures and tools to facilitate the processes of finding necessary information. In this case, the search for information on the Internet is used by various professions, but most often by those whose professional functions directly include the processing of information. Such professions include translators.

The Internet's information resources are now the most in demand, and their volume continues to grow. According to the latest figures, new web pages are created almost every three seconds. Of course, not all of them are useful, because there is also so-called "information noise". At the same time, the Internet as an information environment has many advantages that cannot be overestimated: it contains a large amount of information of various types; the information available is constantly updated; information is updated quickly and dynamically; information is available in various languages and data formats; information can be accessed quickly.

3L-Edu 2022: 2nd International Conference on New Trends in Linguistics, Literature and Language Education, May 18, 2022, Kryvyi Rih, Ukraine

✉ svetlanaamelina@ukr.net (S. M. Amelina); tar@nubip.edu.ua (R. O. Tarasenko); vashyn@nubip.edu.ua (V. D. Shynkaruk)

🌐 <https://nubip.edu.ua/node/6246> (S. M. Amelina); <https://nubip.edu.ua/node/8840> (R. O. Tarasenko); <https://nubip.edu.ua/node/95553> (V. D. Shynkaruk)

🆔 0000-0002-6008-3122 (S. M. Amelina); 0000-0001-6258-2921 (R. O. Tarasenko); 0000-0001-8589-4995 (V. D. Shynkaruk)



© Copyright for this paper by its authors, published by Academy of Cognitive and Natural Sciences (ACNS). This is an Open Access article distributed under the terms of the Creative Commons License Attribution 4.0 International (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



ACNS Conference Series: Social Sciences and Humanities

It is therefore relevant to facilitate the translator's work to study the available tools for searching and processing the information needed [1], as there are now various options, ways and systems available for the user to carry out information searches. Information retrieval in digital resources can be carried out, *inter alia*, via internet directories, internet search engines, internet encyclopaedias and similar information portals, and semantic search engines.

The purpose of this paper is to review how Internet search engines can be used to support the information needs of a translator.

2. Literature review

Internet search engines have already attracted the attention of researchers over the past two decades, as evidenced by a number of relevant publications describing different types of Internet searches, comparing available search engines, studying the effectiveness of information searches, etc. [4, 7, 9, 16].

In particular, Griesbaum, Rittberger and Bekavac [8] have published the results of a comparative analysis of the German search engines AltaVista.de, Fireball.de, Google.de and Lycos.de.

All researchers agree that Google is the most popular search engine. Dirscherl [5] states this for European users in particular: "When Europeans search for something on the internet, they usually do it with Google. The Google search engine is way ahead of competitors like Yahoo! or MSN". At the same time, while not denying this view, Gil [7] notes: "Google might be the biggest but there are other, arguably better, search engines". We would agree with this position, as the translator's awareness of the availability of a number of different search engines allows to optimise their search, taking into account the options of these engines.

Recently there has been an increase in reports of new search engines and publications on their use. In particular, Weiß [15] discusses the search engine Brave Search, focusing on its user privacy and advertising filtering.

Despite the increasing attention of both researchers and practitioners to the study and use of Internet search engines, the issue of their use in the professional activities of translators has not yet been addressed. Therefore, it requires a separate study and, subsequently, the development of practical recommendations.

3. Result and discussion

3.1. Information retrieval by translators

When examining the characteristics of translators' activities at the present stage, researchers pay attention to their use of auxiliary aids in the process of preparing and performing translation [2, 10, 12, 13]. This also applies to the provision of information needs of the translator.

In order to guide translators through the available auxiliary resources, the European Commission has set up a special website [14], which demonstrates the attention European institutions are paying to improving the information competence of translators.

The European Masters in Translation programme stipulates the following skills within the technological competence of a translator: "Make effective use of search engines, corpus-based

tools, text analysis tools and CAT tools” [11].

Information retrieval consists of several steps to obtain information that is important for further processing or using it to clarify, detail, and verify already existing information of the text that is being translated. The first thing to do is to clearly define the information need and therefore formulate the information request correctly; select the search engine to be used for the information search and focus on specific information sources. The next step is to obtain information on the formulated information request and analyse it in terms of relevance and usability.

3.2. Possibilities for translators to use some search engines

Our aim is not to compare the available Internet search engines, but only to determine how they can be used in the work of translators.

It is advisable to familiarise with the main search engines that can be useful for translators to obtain reference, additional or clarifying information when translating.

The most popular, by far, is the Google search engine, recognised as the world leader in Internet searches. It has a market share of about 80 per cent.

Google’s search engine is based on the implementation of computer programs called Webcrawlers. They automatically search the internet by analysing web pages. The links found are the basis for the inclusion of pages in the ever-growing search index, even though its volume is now very large – it already exceeds three billion URLs. Certainly, this amount of information can satisfy a wide range of information needs of translators.

Found web pages are categorised by keywords and concepts, so users receive ranked search results. The PageRank is calculated based on more than 200 factors, which include primarily relevance. The said indicator displays the link popularity, that is, the number and quality of links directed to the web page in question. The more frequently a page is linked, the higher its link popularity, and therefore the higher the relevance of the web page is judged, which in turn increases the number of its priority hits in search results.

As well as searching the web, Google also offers many other information search services, such as full-text searches for PDF or Microsoft Word documents, image searches, news searches and research paper searches via Google Scholar.

In addition, Gmail, the video portal YouTube, the automatic website translation feature and at least the free Android and ChromeOS operating systems, as well as the Google Chrome Internet browser, are also in operation.

At the same time, there are a number of alternative search engines that are less popular than Google but are quite commonly used in some countries. Being familiar with such search engines and appreciating their possibilities for use in professional activities is useful precisely for the profession of translator. After all, a translator is always an intermediary between different languages and therefore “collaborates” with different countries. It is therefore important to know and be able to use the resources that are more or less prevalent in these countries.

One of the most widely used search engines in the world is Yahoo (figure 1). It is predominantly used in the United States of America, but more and more users from other countries are beginning to use the service because it is convenient enough to conduct searches. The extensive list of languages, which currently includes 38 language versions, also contributes to its popularity.

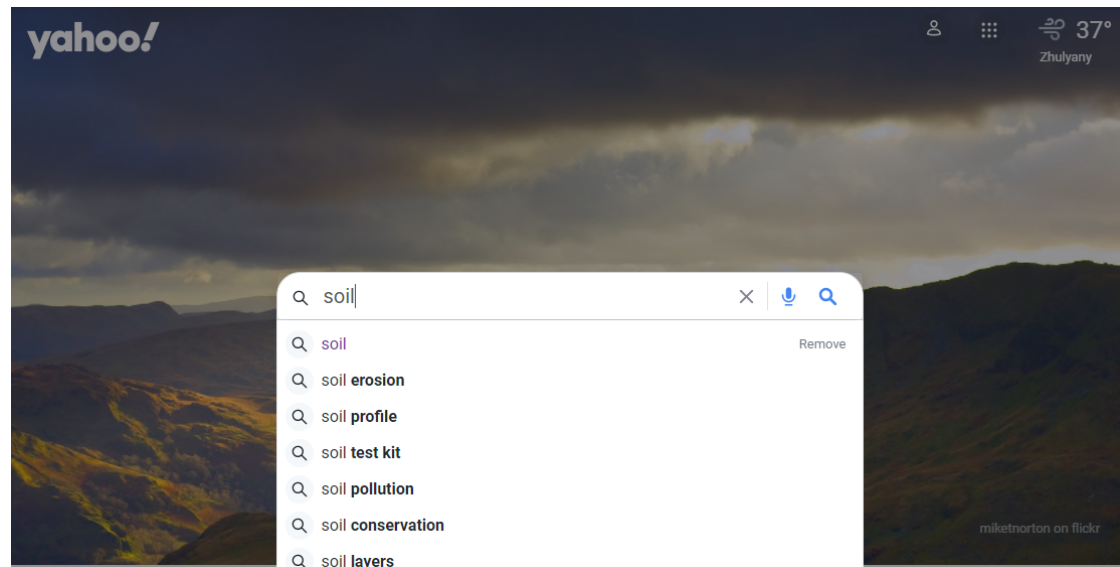


Figure 1: Search for the term “Soil” on Yahoo search engine.

The most well-known search engine emphasising compliance with data protection is the American search engine DuckDuckGo. The number of its users is growing rapidly and steadily, primarily due to the provider’s promise not to collect or store any user data. DuckDuckGo (figure 2) makes it possible to search the Internet not only for websites but also for photos and videos.



Figure 2: DuckDuckGo search engine page.

The Netherlands search engine Ixquick, which developed the Startpage search engine (figure 3), like the US search engine DuckDuckGo, also provides assurances that it does not store any user data, which includes the IP address, browser and the search query itself. Only the number of daily queries is stored. This meta-search engine allows you to search for photos and videos as well as phone book searches. It has been noted that this search engine allows users to retrieve Google search results while protecting user privacy by not storing personal information or search data. In addition, the benefits of this search engine can also include the fact that it reports potentially dangerous sites.

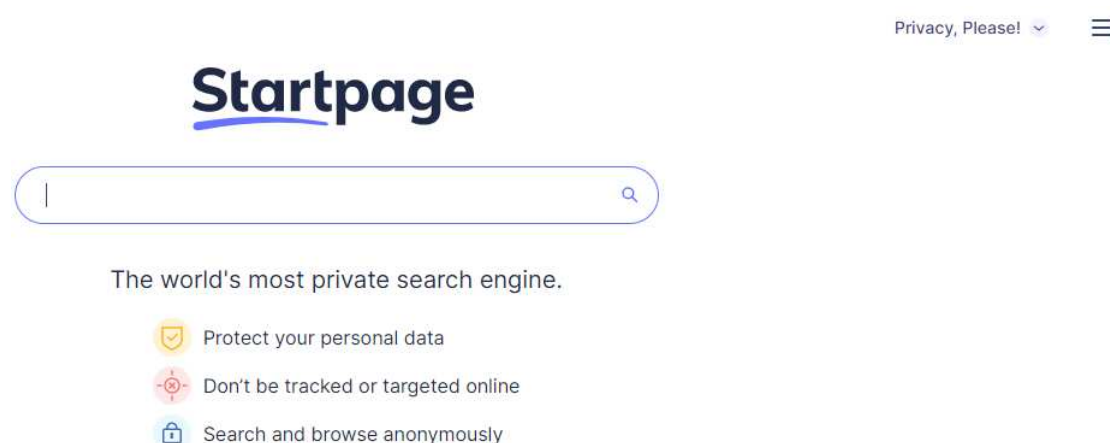


Figure 3: Startpage system search page.

Another well-known so-called meta-search engine is the German search engine MetaGer (figure 4). It automatically scans the entire range of web search engines and then summarises the results. The user can also decide which sources should be used for the search. In addition to the usual web search, it is possible to search for photos, news, and scientific papers with this system. MetaGer does not collect IP addresses or store visitors' cookies.



Figure 4: MetaGer system search page.

It is worth noting that one of the great advantages of the MetaGer search engine is that it is a meta-search engine that combines the search results of several machines, evaluates them according to its own criteria, and therefore gives different results. In some cases, it is this presentation of information search results that can satisfy a translator's needs. Using

a metasearch engine with a search procedure makes it easier for the translator to find the information they need, as they only have to formulate one search query. MetaGer blocks certain web pages that the system regards as questionable sites or sites with incorrect information, poor quality, etc.

Another common and potent German meta-search engine is Wegtam Search Agent (figure 5). According to the provider's own data, its aim is to focus on relevant information and avoid commercial influences. The system also respects privacy provisions and promises not to store any personal information. In addition to web search, there are separate search options for news, people, and videos.

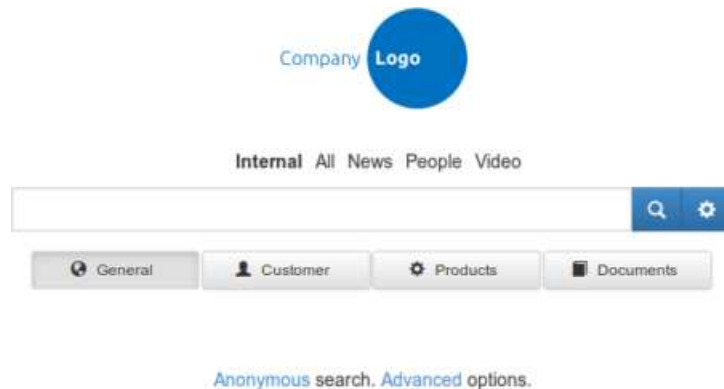


Figure 5: Wegtam Search for Business sample page.

Cybersecurity experts recommend using the European search engine Exalead (figure 6). Like most others, it also allows for searching images, videos or articles in the Wikipedia online encyclopedia, in addition to web searches. When conducting an information search on Wikipedia, the user can specify one of the languages in which material is to be retrieved – English, German, French, Dutch, Italian, Spanish.

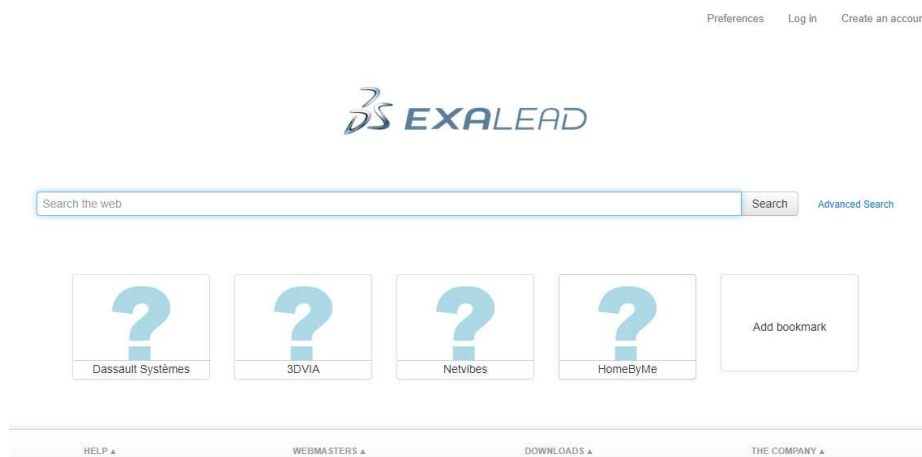


Figure 6: Exalead search engine page.

In 2013, the French search engine Qwant was officially launched (figure 7). Its developers claim that they have a particularly strict privacy policy. Instead of Google Analytics, the provider uses Piwik, which is the only analytical tool certified by French data protection authorities [6]. In addition to normal web searches, Qwant also offers media searches and people searches. Web searches display news, social media content and topical videos on the same page. However, users feel that this presentation of a mix of different materials together makes it difficult to perceive the results of an information search.

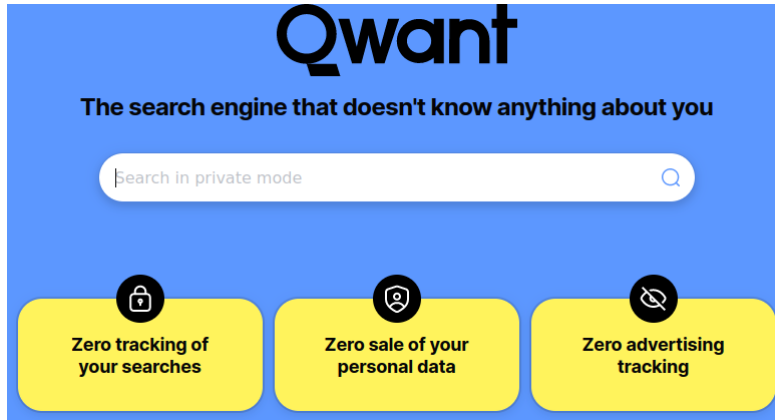


Figure 7: Qwant system search page.

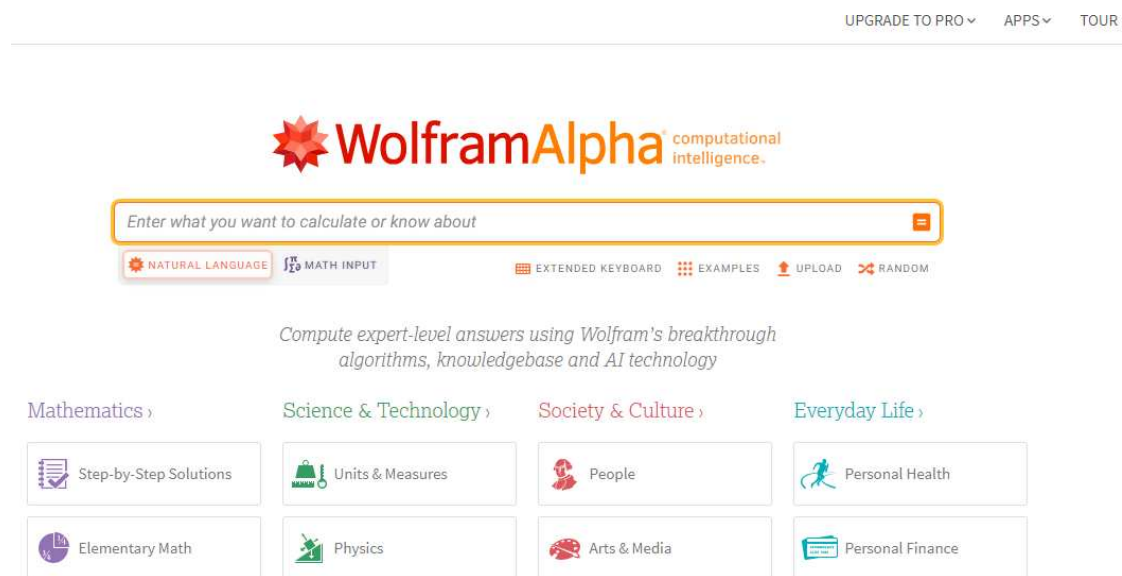


Figure 8: Wolfram|Alpha system search page.

The Wolfram|Alpha search engine (figure 8) differs significantly from traditional search engines such as Google or Bing. Instead of directing users to the most relevant websites, the

system tries to process information based on algorithms and provide the user with direct answers to their questions.

The basic approach of Wolfram|Alpha is that it is only possible to get an answer to a question for which an exact answer is possible. For example, if the query concerns the date of birth of a celebrity or a historically famous person, the correct answer can be obtained through Wolfram|Alpha. The developers of Wolfram|Alpha are guided by the fact that this system works with facts, but does not give estimates [3].

The same principle of “question and answer” applies to searching for information in the Ask system (figure 9).

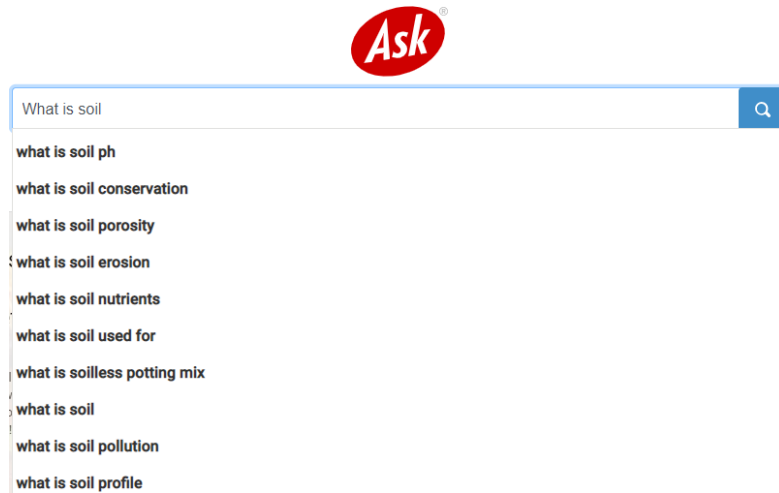


Figure 9: Search for information on the essence of “Soil” in the Ask search engine.

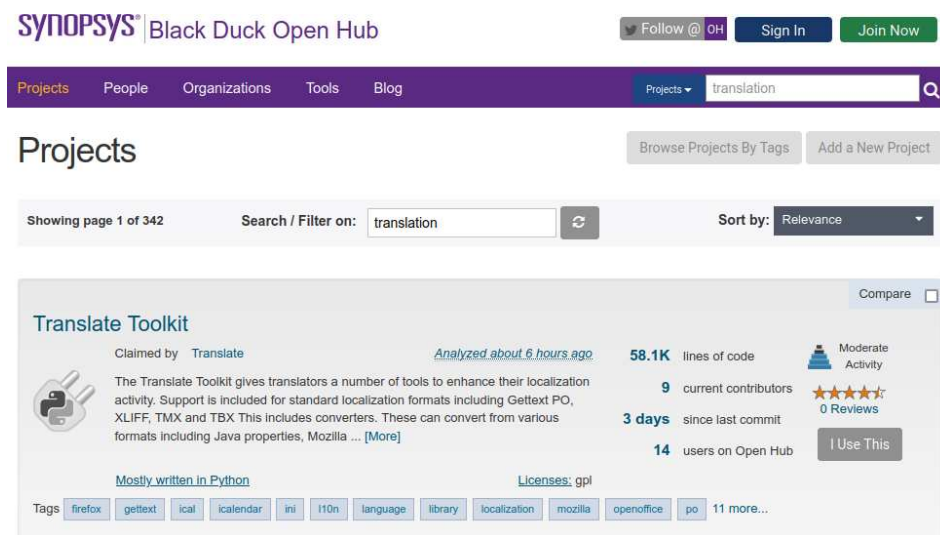


Figure 10: Black Duck Open Hub search results.

Specialised search engines deserve special attention. These include Ohloh, which is a directory of open-source projects. In 2017, the system was renamed to Black Duck Open Hub (figure 10). This search engine searches for a given keyword. Users can target searches by class, function or method, and the results can be sorted by project, language or file extension.



Figure 11: WorldWideScience system search page.

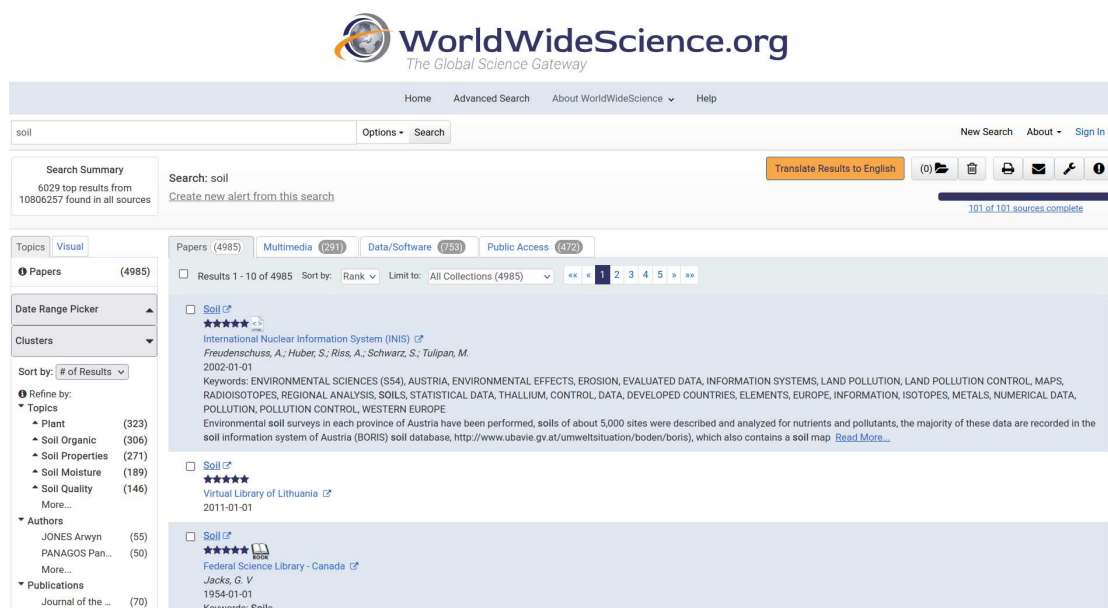


Figure 12: Search results for “Soil” in the WorldWideScience search engine.

WorldWideScience (figure 11, figure 12) is a search engine for scientific texts. The results can be sorted by author, country of origin and text type. In addition to scientific articles, the search engine can also find relevant multimedia content and data files. The site allows more than 200 million documents to be searched. The US-based Office of Scientific and Technical Information,

which collaborates with various organisations around the world, operate the system.

The WorldWideScience system is particularly important for translators who translate scientific and technical texts. By combining national and international scientific databases and portals, it is possible to search for relevant information in databases from all over the world in real time (figure 13).

WorldWideScience.org
The Global Science Gateway

Home Advanced Search About WorldWideScience Help

Enter Search Terms below, then select Language and Databases.

Full Record
Title
Author
Date Range
From To
Limit to Available Full-Text
Preferred Language
Arabic Deutsch English Español Français 日本語 한국어 Português Русский 中文
Translations powered by Google translator
Clear Search AMPLYFI

All Categories
English Sources
African Journals Online (AJOL)
Agris
ArXiv.org
ASEP - Repository of the Czech Academy of Sciences
British Library Electronic Table of Contents (United Kingdom)
CDC Stacks
CERN Document Server
CiNii (Japan)
Multilingual Sources
BazTech (Poland)
Chinese Academy of Sciences Institutional Repositories Grid
German National Library of Science and Technology (TIB) (German)
HAL Archives (French)
Institute of Scientific and Technical Information of China (Chinese)
Multimedia Sources
Centers for Disease Control (CDC) Podcasts
CERN Multimedia
Medline Plus
NASA Video Gallery
National Science Foundation
ScienceCinema
TIB AV-Portal
Data and Software Sources
Australian Antarctic Data Centre

Figure 13: Search options for different queries in WorldWideScience.

An interesting option is the Million Short search engine (figure 14, figure 15). In general, it works on the same principle as the other search engines. However, unlike Google or Bing, it also allows you to exclude the top 10, 100, 1,000, 10,000, 100,000 or a million results, if desired or required. This compares to starting a search on, for example, page 100 of Google's search

million short BETA

What haven't you found? Don't remove any sites Settings

Figure 14: Million Short system search page.

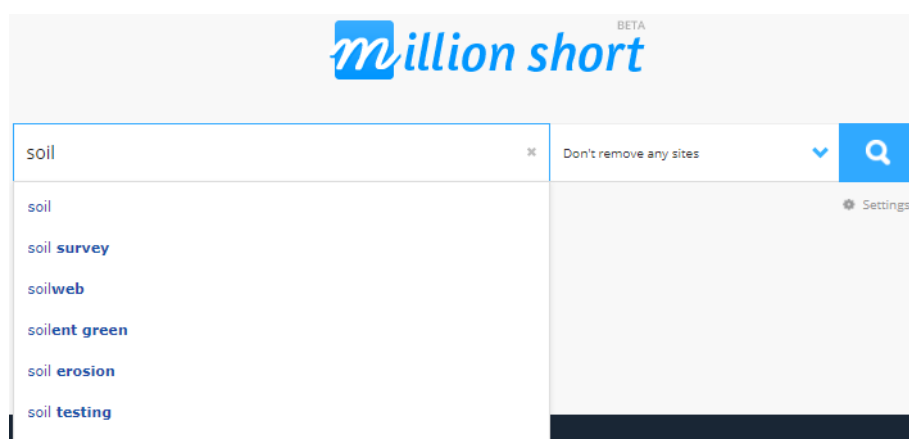


Figure 15: Search results for “Soil” in the Million Short search engine.

results. Some users find this feature quite practical. After all, Million Short can in principle be used in the same way as any other search engine, without applying the extra option offered.

In the last few years, a number of new search engines have emerged with positive user feedback, both in terms of search efficiency and privacy. These are, for example, Searx (figure 16) and Search Encrypt (figure 17). These systems also offer some additional benefits. In particular, it is possible to optimise searches using Searx by limiting results in many ways – by language, file type, location, age of results, etc.

The above examples of search engines that translators can use to support their information needs are certainly not an exhaustive list.



Figure 16: SearX system search page.



Figure 17: Search Encrypt system search page.

4. Conclusions

In order to carry out an effective information search, it is advisable for the translator to clearly define their information needs and formulate their query clearly. Depending on whether general or specific information needs to be found and other query criteria, it is possible to save time by using search engines that operate according to different principles of information retrieval among Internet resources. In particular, it may be a keyword search, a subject search, etc., or a metadata search.

It is important for the translator, as an intermediary between multilingual communicators, to be able to use search resources that are common among speakers of these languages. This will allow for a more accurate and comprehensible translation for speakers of the target language.

In addition, in the context of cyber threats, the modern translator should also consider, when choosing a search engine, its privacy policy, which avoids the loss of personal data. On the other hand, a useful function of some search engines is to warn about dubious and dangerous websites appearing in search.

References

- [1] Amelina, S.M. and Tarasenko, R.O., 2018. The essence of the stages of forming the information competence of the translator. *Information Technologies and Learning Tools*, 67(5), pp.44–55. Available from: <https://doi.org/10.33407/itlt.v67i5.2276>.
- [2] Amelina, S.M., Tarasenko, R.O., Semerikov, S.O. and Shen, L., 2022. Using mobile applications with augmented reality elements in the self-study process of prospective translators. *Educational Technology Quarterly*, 2022(4), p.263–275. Available from: <https://doi.org/10.55056/etq.51>.
- [3] Bas, S.V. and Slovak, K., 2014. Ways of processing requests and characteristics of mobile access to Wolfram|Alpha. *CTE Workshop Proceedings*, 2, p.309–316. Available from: <https://doi.org/10.55056/cte.226>.
- [4] Deka, S.K. and Lahkar, N., 2010. Performance evaluation and comparison of the five most used search engines in retrieving web resources. *Online Information Review*, 34(5), pp.757–771. Available from: <https://doi.org/10.1108/14684521011084609>.
- [5] Dirscherl, H.C. Europäer bevorzugen klar Google. Available from: <https://www.pcwelt.de/news/Suche-Europaeer-bevorzugen-klar-Google-67041.html>.
- [6] Fuest, B., 2014. Qwant: Europäische Suchmaschine verspricht Datenschutz. Available from: <https://www.welt.de/wirtschaft/webwelt/article125423712/Europaeische-Suchmaschine-verspricht-Datenschutz.html>.
- [7] Gil, P., 2023. The Best Search Engines of 2023. Available from: <https://www.lifewire.com/best-search-engines-2483352>.
- [8] Griesbaum, J., Rittberger, M. and Bekavac, B., 2002. Deutsche Suchmaschinen im Vergleich: AltaVista.de, Fireball.de, Google.de und Lycos.de. In: R. Hammwöhner, C. Wolff and C. Womser-Hacker, eds. *Information und Mobilität. Optimierung und Vermeidung von Mobilität durch Information. Proceedings des 8. Internationalen Symposiums für Informationswissenschaft*. Konstanz: UVK, *Schriften zur Informationswissenschaft*, vol. 40,

- pp.201–223. Available from: http://www.dipfdocs.de/volltexte/2020/18249/pdf/Deutsche_Suchmaschinen_im_Vergleich_A.pdf.
- [9] Lewandowski, D., 2015. Evaluating the retrieval effectiveness of web search engines using a representative query sample. *Journal of the Association for Information Science and Technology*, 66(9), pp.1763–1775. Available from: <https://doi.org/10.1002/asi.23304>.
- [10] Manning, C.D., Raghavan, P. and Schütze, H., 2009. *Introduction to Information Retrieval*. Cambridge, England: Cambridge University Press. Available from: <https://nlp.stanford.edu/IR-book/information-retrieval-book.html>.
- [11] Scudamore-Trezek, J., 2021. European Masters in Translation. Available from: <https://www.germantranslationtips.com/European-Masters-in-Translation.html>.
- [12] Tarasenko, R., Amelina, S., Semerikov, S. and Shen, L., 2022. Creating a cloud-based translator training environment using Memsources. *Educational Technology Quarterly*, 2022(3), p.203–215. Available from: <https://doi.org/10.55056/etq.33>.
- [13] Tarasenko, R.O., Amelina, S.M. and Azaryan, A.A., 2020. Improving the content of training future translators in the aspect of studying modern CAT tools. *CTE Workshop Proceedings*, 7, pp.360–375. Available from: <https://doi.org/10.55056/cte.365>.
- [14] Translation and drafting resources, 2016. Available from: https://ec.europa.eu/info/resources-partners/translation-and-drafting-resources_en.
- [15] Weiß, E.M., 2021. Brave Search: Browser veröffentlicht Suchmaschine als Public-Beta. Available from: <https://www.heise.de/news/Brave-Search-Browser-veroeffentlicht-Suchmaschine-als-Public-Beta-6115843.html>.
- [16] Wolff, C., 2000. Vergleichende Evaluierung von Such- und Metasuchmaschinen im World Wide Web. In: G. Knorz and R. Kuhlen, eds. *Informationskompetenz – Basiskompetenz in der Informationsgesellschaft: proceedings des 7. Internationalen Symposiums für Informationswissenschaft, Darmstadt, 8-10 November 2000*. Konstanz: UVK, *Schriften zur Informationswissenschaft*, vol. 38, pp.31–48. Available from: <https://epub.uni-regensburg.de/6803/>.