

SEO - Oriented Bulk WEB Parser

Pavel Korniienko, PhD Vladislav Khotunov*¹, Igors Babics²

¹*Cherkasy State Business-College*

²*ISMA University of Applied Sciences, Riga, Latvia*

*Corresponding author's email: vkhotunov@gmail.com

Abstract

Today, online translators have gained their popularity, including thanks to the rich choice of languages - despite the fact that people over time are becoming more educated, and learn more languages, to have at an important moment from a professional or household point of view, an application capable of translating a few lines into the desired language in two clicks. Even English, the most widely spoken language in the world, sometimes requires such programs. The reason that translators are mostly created as online resources is that they have sophisticated word gluing algorithms to translate complete sentences, and have an extensive database that, in the years when translators first appeared, was simply impossible for the average person to store because of too little physical memory of the speakers. Since then, online translators are still popular, as the Internet is used by most people in the world, which means that a translator will always be conveniently available for use. Also, translators are popular in professional environments - including SEO. SEO - this is the direction, or rather the process of placing sites in the top of search engines to gain popularity.

Keywords: SEO structure, Translator, Parser, WEB development

1 Object of study

The object of the study was a resource capable of optimizing translation systems to simplify thematic tasks in SEO structures, in particular, to translate multiple texts at a time. One of the ways to promote the site and output to the top search engines is to create a grid of sites aimed at disseminating information about the project. One of the stages of creating such sites is filling it with authentic content, as it directly affects its effectiveness. Also, along with authenticity, it is necessary to monitor the constructive component of the content, despite its relevance to the topic. There are several ways to search for content for such purposes, one of which is to search for sites that fill the relevant topics in another language. This method is one of the most effective, as translated texts are more likely to be unique to search engines, thus increasing their relevance. This means that it makes sense to deal with its practicality for quick implementation.

2 Subject of study

This project is focused on the processing of found content on sites, in particular translation and layout in a convenient format for easy processing and export to the necessary services for further processing. The package translator is

aimed at managers - specialists involved in creating meshes and filling sites with content, it simplifies the work on translation. Under normal conditions, each article required for the project should be translated individually and additionally checked for syntax errors. This project will accelerate this process during the translation phase. It will be possible to translate several articles at once by inserting several links into the field and all articles from the list will be translated. You can also choose a tariff - Google translator or DeepL - each of them has its advantages and disadvantages. The translated articles will be conveniently collected in archive from where it will be possible to check up on uniqueness and to take in Google Docs for the further check on syntactic errors, and also publications on the domain by means of CMS chosen by the user. Content on sites is located in HTML layout format - the page elements themselves are located in tags. The main tags in which the necessary materials are located - the tag "<p>" - in other words, "paragraph", as well as tag headings "<H1> - <H5>", they are all located in a peculiar hierarchy. The project has parser properties - the script will pull the text out of the specified page and translate it in the translator, depending on which one you choose, and you will get the translated text. Texts in the archive, each text will be saved in a ".docx" document.

3 Project counterparts and their comparison

Advantages of search engines over the project:

ADVANTAGES					
My project	Google translate	deftpdf.com	radugaslov.ru	deepl.com	onlinedoctranslator.com
Translation of text on web resources	Many languages	Translates files, "docx", or "pdf".	Reverse translation	Better translation quality	Translates all basic formats of Microsoft Word
Several translation databases	Voicing translation	Allows saving files in your online account for future use	Voicing translation	PC Application	
Easy choice of text structure	Offline application				

Disadvantages of the search engine counterparts on the project:

DISADVANTAGES					
My project	Google translate	deftpdf.com	radugaslov.ru	deepl.com	onlinedoctranslator.com
Unsuitable for translations of several words or sentences.	Machine translation	Unsuitable for translations of several words or sentences.	Few translation languages available	Few translation languages available	Unsuitable for translations of several words or sentences.
Not suitable for offline document translation					

4 Conclusion

This development will allow to use translation more

effectively in the professional field - SEO, thanks to which it will be possible to quickly implement the tasks associated with the translation of texts.