



Site search for multi-language sentence translations



Company

[Tatoeba.org](https://tatoeba.org) is a large database of sentences and translations. Its content is ever-growing and results from the voluntary contributions of thousands of members.

Tatoeba provides a tool for you to see examples of how words are used in the context of a sentence. You specify words that interest you, and it returns sentences containing these words with their translations in the desired languages. The name Tatoeba (for example in Japanese) captures this concept.

The project was founded by Trang Ho in 2006, hosted on Sourceforge under the codename of multilangdict.



Challenge

A long time ago (2010) we were using Lucene and decided to switch to Sphinx due to memory restrictions. Before switching to Manticore we had a quick look at other solutions like ElasticSearch, but rewriting all the search-related code would be a big effort. While Elastic has a lot of fancy stuff, our data is pretty “flat” (sentences with metadata), and Manticore just fits in.

We knew about Manticore since November 2017 but took us a while to effectively migrate. We were using Sphinx but lately was crashing quite often and as a result, making our homepage completely broken ([#1767](#)).



Solutions and results

- **Migrating** from Sphinx to Manticore and as a result:
 - **Search daemon** does not block anymore while it was happening with Sphinx.
 - **Improved search speed** and as a result the speed of the whole website with 220K-280K searches per month or 7.5-10K per day.

Trang Ho & Gilles Bedel

[Tatoeba](https://tatoeba.org)



Contact or visit us if you have any questions.

E-mail: contact@manticoresearch.com

Services page: <https://manticoresearch.com/services/>