

Cross-Lingual Retrieval

March, 2008

Abstract

Cross-lingual retrieval is the technology that retrieves information from various documents and databases written in languages different from those in which key words are written. For example, it could be used to retrieve information from English academic papers or Chinese news articles using Japanese key words.

Mainly, there are two methods of cross-lingual retrieving;

- (1) To retrieve information from documents and databases after translating key words into the same language as those in which the documents and databases are written.
- (2) To translate documents and databases into the same language as those in which key words are written prior to the retrieval.

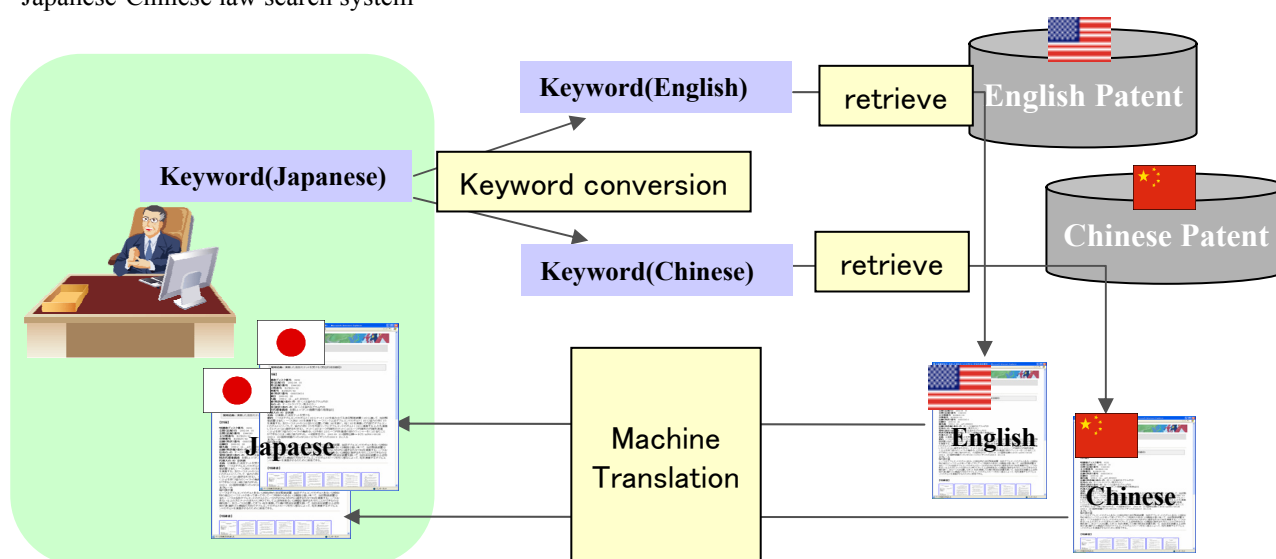
Method (1) would be the most appropriate for retrieving information in various languages. However, in that case it would be necessary to assist users in their understanding, for example, by machine-translating the outcome into the language they use.

Technology

It is important to obtain clear and precise keywords without ambiguities in their conversion process. It is also important to choose the most appropriate key words from various choices in their conversion process. For this purpose we have developed technologies for translating keywords in context. More accurate cross-lingual retrieval will be achieved through technologies such as making meaningless words invalid key words- such as “it”, “do” – by selecting the most appropriate key words using machine translation technologies and weighing key words based on their frequency of appearance in the documents and databases.

Application Examples

- ATLAS V14
<http://www.fujitsu.com/global/services/software/translation/atlas/>
- Accela BizLingo V3
<http://www.fujitsu.com/global/services/software/translation/bizlingo/>
- Japanese-Chinese law search system



Example of cross-lingual patent database retrieval system