OASIS at NTCIR-6: On-line Query Translation for Chinese-Japanese Cross-Lingual Information Retrieval

Vitaly Klyuev
University of Aizu
Tsuruga, Ikki-Machi, Aizu-Wakamatsu, Fukushima, 965-8580, Japan
vkluev@u-aizu.ac.jp

Abstract

This paper reports results of Chinese – Japanese CLIR experiments using on-line query translation techniques. Approaches to employ English as a pivot language and to utilize several on-line translation systems are introduced. They were tested on NTCIR – 3, 4, 5, and 6 collections. Proposed procedures can be helpful under certain circumstances.

1   On-line translation techniques

We worked with different on-line translation systems. Techniques which do not require utilizing training data and adjusting the parameters of the retrieval system were applied. Our strategy was as follows:

o   Apply on-line systems to translate queries directly from Chinese into Japanese,
o   Utilize the model of Chinese – English – Japanese translation, where English is the pivot language,
o   Merge the translation results produced by different systems for the possible expansion of queries.

We employed Mecab as a segmentation tool. The vector space model was the basis of our search engine. We used information about the part of speech of the words generated by the morphological analyzer: Nouns and verbs were filtered in the "D" runs.

Table 1. Results of experiments

Run	R-Precision	Precision at 5 docs	Precision at 10 docs	Comments
OASIS-C-J-T-03 (WorldLingo, Excite)	0.13	0.20	0.18	Merging results of two on-line translation systems, (Stage 1, "Relaxed" relevance judgment)
Babel Fish	0.0955	0.1600	0.1300	English as a pivot language (Stage 1, "Relaxed" relevance judgment, "T" runs)
DictDotCom	0.1112	0.1760	0.1620	
Google	0.1163	0.1840	0.1680	
Google – Babel Fish	0.1087	0.1840	0.1640	
Babel Fish – Google	0.1068	0.1800	0.1620	
DictDotCom – Babel Fish	0.1008	0.1720	0.1480	
Google – DictDotCom	0.1139	0.1680	0.1620	
OASIS-C-J-D-01-N3 (WorldLingo, Excite)	0.1336	0.1952	0.1643	Merging results of two on-line translation systems, (Stage 2, "Rigid" relevance judgment)
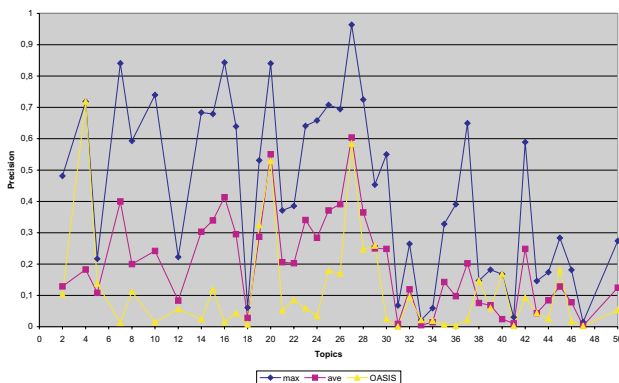
Figure 1. Compared precision Stage 2: MAP_C-J-D-Rigid – OASIS-C-J-D-N3

2   Conclusions

The efficiency of on-line translation systems were tested when they were applied to translate queries from Chinese into Japanese for the CLIR task.

Results showed that outcomes of translation systems are different because the systems use various approaches, different dictionaries, etc.

Using several on-line translation systems and combining their translation results may be helpful to expand queries and improve the quality of retrieval for the CLIR task. This strategy can be used as part of a set of measures to make results of cross-lingual retrieval more accurate. Our experiments to employ English as the pivot language did not produce the significant changes in the retrieval.