

Bilingual dictionary approach for Malay-English cross-language information retrieval

ABSTRACT

Cross-language information retrieval (CLIR) is the process of providing queries in one language and returning documents relevant to that query which is written in a different language. A popular approach to CLIR is to translate the query into the language of the documents being retrieved. One of the simplest and most effective methods for query translation is to perform dictionary look up based on a bilingual dictionary. Direct translation using bilingual dictionary prune three main problems: (1) knowing how a term expressed in one language might be written in another; (2) deciding which of the possible translations should be retained and (3) deciding how to properly weight the importance of translation alternatives when more than one is retained. We evaluated the effectiveness of Malay-English CLIR system using bilingual dictionary approach. In this study, we presented the evaluation results for dictionary-based CLIR. A document collection containing newspaper articles and a related set of 35 search queries were used in this test. First, monolingual baseline queries were created manually in Malay and English languages. Secondly, queries in Malay language were automatically translated into English language, and vice versa. There are two basic translation approaches using bilingual dictionary: select the first translation listed in the dictionary and select all translations listed in the dictionary, for each query. Then, alternative weighting scheme were applied to the second query translation approach, select all translations listed in the dictionary, to enhance retrieval performance. These three experiments were evaluated using Mean Average Precision (MAP) and Average Recall-Precision graph. The results were compared to monolingual IR for Malay and English document collection, respectively.

Keyword: Cross-language information retrieval; Bilingual dictionary; Query translation