Query translation using concepts similarity based on Quran ontology for cross-language information retrieval.

Abstract

In Cross-Language Information Retrieval (CLIR) process, the translation effects have a direct impact on the accuracy of follow-up retrieval results. In dictionary-based approach, we are dealing with the words that have more than one meaning which can decrease the retrieval performance if the query translation return an incorrect translations. These issues need to be overcome using efficient technique. In this study we proposed a Cross-Language Information Retrieval (CLIR) method based on domain ontology using Quran concepts for disambiguating translation of the query and to improve the dictionary-based query translation. For experimentation, we use Quran ontology written in English and Malay languages as a bilingual parallelcorpora and Quran concepts as a resource for cross-language query translation along with dictionary-based translation. For evaluation, we measure the performance of three IR systems. IR1 is natural language query IR, IR2 is natural language query CLIR based on dictionary (as a Baseline) and IR3 is the retrieval of this research proposed method using Mean Average Precision (MAP) and average precision at 11 points of recall. The experimental result shows that our proposed method brings significant improvement in retrieval accuracy for English document collections, but deficient for Malay document collections. The proposed CLIR method can obtain query expansion effect and improve retrieval performance in certain language.