Abstract

This paper describes our preliminary research on an understanding-based translingual information retrieval system for which the input to the system is a query sentence in English, and the output of the system is a set of documents either in English or in Korean. The understanding module produces a meaning representation -- called semantic frame -- of the input sentence where the predicate-argument structure and the question-type of the input are identified, and each keyword is assigned its concept category. The translingual search module performs a search on an English and Korean bilingual corpus tagged with concept categories. The results of our preliminary experiment, performed on a document set consisting of slides and notes from English and Korean briefings in a military domain, indicate that an understanding-based approach to information retrieval combined with a concept-based search technique improves both precision and recall compared with a keyword match technique without understanding for both monolingual- and translingual retrieval. Current work is directed at further development of the system, and in preparation for tests on larger corpora.