

Hindi and Telugu to English CLIR using Query Expansion

Prasad Pingali and Vasudeva Varma

Abstract

This paper presents the experiments of Language Technologies Research Centre (LTRC) as part of their participation in CLEF2 2007 Indian language to English ad-hoc cross language document retrieval task. In this paper we discuss our Hindi and Telugu to English CLIR system and the experiments using CLEF 2007 dataset. We used a variant of TFIDF algorithm in combination with a bilingual lexicon for query translation. We also explored the role of a document summary in fielded queries and two different boolean formulations of query translations. We find that a hybrid boolean formulation using a combination of boolean AND and boolean OR operators improves ranking of documents. We also find that simple disjunctive combination of translated query keywords results in maximum recall.