Query translation architecture for Malay-English cross-language information retrieval system

ABSTRACT

This paper discusses research on query translation events in Malay-English Cross-Language Information Retrieval (CLIR) system. We assume that by improving query translation accuracy, we can improve the information retrieval performance. The dictionary-based CLIR system facing three main problems: translation ambiguity; compound and phrase handling and proper names translation. The use of natural language processing (NLP) techniques, such as stemming, Part-of-Speech (POS) tagging is useful in query translation process. Hence, n-gram matching technique has successfully applied to information retrieval (IR) system for phrases and proper names translation. The proposed query translation architecture consist of stemming, Part-of-Speech (POS) tagging and n-gram matching techniques is useful in CLIR system as well as search engine application.