AN ONTOLOGY BASED AUTOMATIC ANNOTATION AND SEMANTIC INFORMATION RETRIEVAL FROM TAMIL DOCUMENTS

ABIRAMI. S AND MANJULA. D

Abstract

The use of Ontologies to overcome the limitations of keyword-based search has been put forward as one of the motivations of the Semantic IR. We propose a model, which includes an ontology-based scheme for the automatic annotation of Tamil documents and a retrieval system. The retrieval model is based on an adaptation of the classic vector-space model, including an annotation weighting algorithm, and a ranking algorithm. Keyword search includes preprocessing like tokenizing, stop word removal, Root word extraction, noun & verb separation and frequency occurrence from Tamil documents whereas Ontology based Automatic annotation of documents is like a Knowledge Management portal, done through the exploitation of presentation structures embedded within documents. The approach introduced is fairly simple yet highly effective in annotating documents automatically. Semantic search is combined with conventional keyword-based retrieval to achieve tolerance to knowledge base incompleteness. Experiments are shown where our approach shows clear improvements with respect to keyword-based search.