Enhanced Arabic Semantic Information Retrieval System Based on Arabic Text Classification

Authors : A. Elsehemy, M. Abdeen , T. Nazmy

Abstract : Since the appearance of the Semantic web, many semantic search techniques and models were proposed to exploit the information in ontology to enhance the traditional keyword-based search. Many advances were made in languages such as English, German, French and Spanish. However, other languages such as Arabic are not fully supported yet. In this paper we present a framework for ontology based information retrieval for Arabic language. Our system consists of four main modules, namely query parser, indexer, search and a ranking module. Our approach includes building a semantic index by linking ontology concepts to documents, including an annotation weight for each link, to be used in ranking the results. We also augmented the framework with an automatic document categorizer, which enhances the overall document ranking. We have built three Arabic domain ontologies: Sports, Economic and Politics as example for the Arabic language. We built a knowledge base that consists of 79 classes and more than 1456 instances. The system is evaluated using the precision and recall metrics. We have done many retrieval operations on a sample of 40,316 documents with a size 320 MB of pure text. The results show that the semantic search enhanced with text classification gives better performance results than the system without classification.