ABSTRACT

Searching the Web has become more challenging due to the rapid growth in information. The documents contain much valuable knowledge about a particular domain. The ontology can be used as main resource to understand the textual information contained within the documents. Ontology based information retrieval matches the relevance of a user generated query against a knowledge-base. The motivation of this thesis is that the terms in the document have multiple meanings. Thus, providing ontology similarity certainly helps to formulate more effective retrieval according to the user's needs. The objective of the research is to define a model for the annotation and retrieval using optimization techniques. This research shows how to apply ontology based annotation method to improve the retrieval. The quality of the solution obtained can be improved by using annotated weights and optimized clustering algorithm.

Another objective is to extract relevant concept from the corpus. The ontology population generates the new instance and results in the semantic annotation of document. The algorithm infers the semantic relatedness between the documents. The main goal of information extraction is to retrieve the relevant information from the document and to create an instance of ontology. Most of the content in the web are in the form of raw texts and they are not semantically annotated. This research aims to use

ontology in information extraction from unstructured resources. The ontologies are very useful in extracting the relevant information and for populating the ontology. Hence, the author has proposed an information extraction technique using ontology. The efficiency of the information extraction process can be increased by annotating documents with semantic information.

The main goal of ontology-driven information retrieval is to enhance search by making use of available semantic annotations and their underlining ontologies. Semantic similarity and indexing focuses on the similarity measure using ontology. It also compares the vector space model with ontology based information retrieval model. The methods are integrated to find the concept relation information, while these concepts are considered to be independent in the term vector space method. The semantic similarity measure is used in calculating the similarity between the concept and documents.

Clustering is an important task in information retrieval to improve the precision and recall. The document clustering can be viewed as a utility to assist in document retrieval. The current keyword methodologies tend to be inconsistent and ineffective when the terms are used for cluster analysis. One major problem with the clustering method is that it does not consider the semantics of the terms. In the traditional retrieval system, keyword method cannot meet the need of user's requirements. To avoid this problem, the ontology similarity model is used in clustering. Optimal cluster based

retrieval model proposes the method of using particle swarm optimization to solve the clustering problems. The objective of the particle swarm optimization clustering algorithm is to discover the proper centroids of clusters for minimizing the intra-cluster distance as well as maximizing the distance between clusters. The performance of the heuristic algorithm is compared with the traditional clustering algorithm. The proposed approach is composed of two main parts. First, the document annotation is done using ontology and second, the optimized clustering approaches to improve the relevance. The novelty of this approach resides in the document annotation and applying particle swarm optimization to retrieve optimized results. The particle swarm optimization clustering algorithm can generate more compact clustering results.

Current information retrieval on the web depends primarily on keywords approaches achieves fairly high recall but at the cost of low precision. The use of ontologies to overcome the limitation of keyword search has been put forward as one of the motivations of the Semantic Web. However, the conceptual model supported by typical ontology may not be sufficient to represent uncertainty information commonly found in many application domains. To handle uncertainty of information and knowledge one possible solution is to incorporate fuzzy theory into ontology.

The Fuzzy Particle Swarm Optimization algorithm is a hybrid method developed in order to combine the properties of fuzzy clustering and Particle Swarm Optimization. The method overcomes the problem of

premature convergence of Fuzzy C-Means algorithm. Fuzzy system for ontology based retrieval suggests method of using ontology and fuzzy to improve the clustering algorithm. The fuzzy clustering may be applied to perform the clustering by combining the meaning of the concepts. The ontology based fuzzy Particle Swarm Optimization clustering algorithm is used to find the optimized clusters. This study investigates the application of fuzzy particle swarm optimization in document clustering. The main objective is to apply the fuzzy particle swarm optimization clustering method on the semantically annotated documents. A fuzzy particle swarm optimization combined with ontology model of clustering knowledge documents is presented and compared to the traditional vector space model. It also overcomes the problems existing in the vector space model commonly used for clustering. The proposed ontology framework provides improved performance and better clustering compared to the traditional vector space model. The increase in F-measure is achieved when ontology as the distance measure in fuzzy particle swarm optimization. The improvement of 11% is achieved by ontology in comparison with keyword search.