ABSTRACT

The World Wide Web has changed the way of communication, business process, search information and entertainment. It has impacted our daily lives. The semantic web is an extension of World Wide Web that can be processed by machines automatically without human intervention. Today variety of applications working towards the syntactic web to semantic web and most interesting web applications is semantic search. Semantic search is an information searching technique in which a search query aims to determine the intent and related meaning of the words a person is used to search. Current search engines provide irrelevant results for user requirements. Semantic search delivers more meaningful search results by evaluating and understanding the search phrase and finding the most relevant results in webpages, databank, server or any other data repository.

This research work proposes semantic web based search techniques to enhance search performance in an information retrieval system. The system has three main techniques: generic semantic search technique, specific semantic search technique and Quality of Service (QoS) based ranking methodology.
Generic semantic search technique is designed for the users who do not have domain knowledge in the area of their search query intentions. Since the users have insufficient knowledge, their search queries may contain vagueness, misrepresentative, matchless kind of information, and wrong search queries. Consequently, the task of this proposed generic search technique is to help the users to disambiguate the queries and match their intentions with search domain knowledge, to precisely find the search information which can best fit their requirements. The generic search technique uses Ontology Knowledge Base (OKB) to store ontology and metadata. Natural Language Processing (NLP) technique is used to spell checker and synonymer for find synonyms. Semantic Case Based Reasoning (SCBR) model is to retrieve efficient information from the OKB.

The specific semantic search technique is designed for the users who have domain knowledge in the area of their search query intentions. The proposed methodology is able to assist them to quickly find required information based on their search queries. In this request, retrieval time is the most important factor for these users. The methodology has three components, namely: ontology knowledge base, Bayesian network and semantic query. The ontology knowledge base is used to store domain specific ontologies and metadata. The Bayesian network is used to estimate the user’s preferable activities and semantic query to quickly retrieve the required results with minimal time.
The QoS ranking methodology is incorporated with the previous results of generic and specific semantic search techniques. The huge retrieved results from OKB may be relevant and irrelevant. A critical problem is occurring to the user, how making the appropriate selection from amongst retrieved metadata. So the problem of re-ranking results has become one of the main problems in the IR field. To fulfil the requirement of users this research using QoS based ranking methodology will search the data semantically and hold the capability to re-rank search results effectively and trying the best to arrange the results which are most relevant for the users. The proposed work defines ranking metric and developed ranking algorithms applied in both generic and specific semantic search in enhancing semantic search performance.

This research has investigated the problem of enhancing search performance in an information retrieval system. In order to increase the precise search results of a generic semantic search, another part of this thesis addresses the quick retrieval of the data which is most important characteristics of information retrieval system. This problem is solved by specific semantic search and QoS based ranking method applied in this research work to improve performance further. The efficiency of the proposed system is validated with educational and tourism data sets. This approach gained improvements in precision, recall, f-measure, mean average precision and time.