Abstract

The World Wide Web (WWW) has evolved as a unanimous powerful information portal which is also a prevalent educational content repository. It allows a learner to retrieve and access available resources (learning content) or documents on any topic or subject. Learners usually take help of existing search engines to find web resources relevant to their need. However, the retrieval of useful and relevant information from the Web requires appropriate query formation by a learner. In addition to it, the resultant web pages from the search are usually very large in number and have varied context. This leaves a learner with the tiresome option to search the required content by visiting each web page one by one. In order to provide relevant web resources to learner, there is a need to improve the existing web search methods by adding context and semantics to the search.

In view of the above challenges, this thesis presents an approach to search relevant web resources based on Concept Ontology, a specially designed hierarchical structure to semantically arrange concepts under a domain. It also explains how these web resources are ranked according to the relevance of concepts, and organized in the knowledge base for effective learning. Ontologies have the potential to construct a dynamic and semantic structure of e-content that can be shared, reused and enhanced for learning purpose. This Concept Ontology forms the basic building block to the work in this thesis. It represents the concepts under a domain which are linked together with the semantic relationships, and the concept can be explored semantically to fetch relevant web pages. The ontology can also be used to determine the semantic relevance of web pages. The structure of the proposed ontology, called Concept Ontology is derived from the Concept Map and SKOS (Simple Knowledge Organization System), which are the best known techniques to represent the conceptual concepts leading to their understanding. They are simpler to build as compared to other expensive ontologies and more expressive to describe relationships among contextual concepts. The Concept Ontology, designed as per the W3C standards is stored as OWL (Web Ontology Language) file, making the resulted knowledge base suitable for the Semantic Web enabled applications. The proposed
approaches such as Concept Term Expansion, FCHC (Focused Crawling based on Human Cognition) and DSRbasedSFC (Dynamic Semantic Relevance based Semantic Focused Crawler) leading to retrieval, and organization of web resources make use of the Concept Ontology.

The traditional search engines use web crawlers to crawl through web pages, documents or resources on the Web, index them and search required resources that match to users’ queries and present them to the user in a ranked form. Such search engines are meant for general use. An ever growing volume of content on the Web remains a challenge for them too. A generic crawl on the Web, for the need of specific web content is very time consuming and requires too many storage resources. Instead, a Focused Web Crawler crawls the Web to reach potentially relevant web pages. Thus fetching only selective web resources, Focused Crawlers need comparatively much lesser crawling time and storage. To accomplish such retrieval needs, Focused Crawlers, FCHC and SFC have been designed, developed and evaluated in the thesis. They gather potentially relevant web resources from different web repositories viz. social bookmarking site and the WWW. Utilization of the semantic knowledge by these proposed focused crawlers during the crawl is their advantageous feature.

The crawled web resources need relevance ranking before presenting to a learner. Ranking algorithms are the core of any information retrieval system as they prioritize web resources based on the predicted relevance. The crawled web resources have been ranked using various attributes (viz. tags, popularity of a web page in the community, individual frequency of semantically related terms to a search topic etc.), gathered during the crawl. The ranks of crawled web resources to a search topic have been computed using social semantic relevance. The relevance metric for predicting ranks of the crawled web resources is based on a well-established Vector Space Model. These potentially relevant web resources are then augmented in the Concept Ontology by linking concepts with various semantic relationships. The knowledge base constructed in this way can be utilized by Semantic Web enabled applications for semantic retrieval.

A system for informal e-Mentoring, a sub-task of e-Learning has been proposed which integrates the sub-processes of retrieval and organization of web
content in the Concept Ontology as a continuous process. In an informal e-Mentoring scenario the conversation between mentor (a knowledgeable person) and mentee (learner) takes place spontaneously, usually for a short duration, till the immediate goal of the mentee is achieved. Also, it does not require any prior investigations on the mentor-mentee relationship. One of the most challenging tasks of informal virtual mentoring includes providing appropriate educational information resources that may interest the mentee. For the purpose of informal e-Mentoring, the system incorporates a Virtual Mentor, a software agent to accomplish the task of providing domain specific information to a mentee, by communicating with other co-operative information gathering agents. Besides virtual mentoring, various components of the system can be used by a number of applications with little or no modifications. This includes semantic retrieval and ranking of tagged resources for any information portal such as digital libraries, bookmarking sites, OER (Open Education Resources) such as Connexions and EdShare. Such a system would be a real help to an information seeker for the all-time availability of mentors and, to a human mentor who would be benefited by saving enough time for other mentoring tasks instead of searching relevant information for the mentee.