7. Conclusion

In this thesis, we investigated the web resources retrieval and their organization in ontology so that the Semantic Web enabled applications may access them through software agents. The primary motive was to develop a system for informal e-Mentoring where a learner can get benefited by the information on a topic of his/her interest. The information on a topic is collated semantically from different sources on the Web.

This chapter points out what has been done and what could be done as future work.

7.1 Summary of the Thesis Contributions

A Multi Agent and Semantic Web based system to impart informal knowledge to a learner is presented in the thesis. The work in this regard, proceeded with an effort to create a structure which could store the information related to various concepts linked conceptually and linked to the relevant content as well. The structure was named as the Concept Ontology. This Concept Ontology was used as a knowledge base for all the software agents and the components of the system.

The system primarily focused on the methods of information retrieval dealing with crawling and organization of the crawled content in the Concept Ontology. This knowledge was utilized by the web crawlers for gathering relevant content from the Web and by the Virtual Mentor (a software agent) to process Mentees’ information needs. In order to achieve these motives, the thesis presented a design and implementation of an Informal E-Mentoring System for knowledge dissemination.

The proposed Concept Ontology (discussed in Chapter 3) consisted of two different ontologies, Domain Ontology (DO) and Content Fragment Ontology (CFO). The design was created with the intent to structurally define the abstract concepts that exist in academic subjects. Thus, the Domain Ontology represented the conceptual knowledge under a domain and, the CFO represented their supportive documents or resources which may help learners to understand the concept of their interests. Design of the Concept Ontology is application and curriculum independent, and therefore the
ontology can be reused, shared and extended for the use by any semantic web enabled application.

The concept terms were expanded using the knowledge from the Concept Ontology. The expansion enabled to include relevant terms related to a concept term, thus made the focused crawl and information retrieval more relevant. Term Expansion (discussed in Chapter 3) was considered as one of the important processes in the whole system and was therefore implemented as a task provided by the OntoAgent and to be consumed by other agents.

The designs of two focused crawlers were proposed and presented in the thesis (Chapter 4) to enable the collection of relevant web resources. A Social Semantic Focused Crawler called FCHC (Focused Crawling based on Human Cognition) was developed using two crawling patterns BFP (Breadth First Pattern) and DFP (Depth First Pattern). Further, DFP was implemented at two different levels, level-1 and level-2. FCHC made use of bookmarked (tagged) web links on social web site and semantic knowledge to prioritize the sequence of web page traversal. The page relevance was computed based on the popularity of the web pages and tags, usually assigned by web users.

The design of a Dynamic Semantic Relevance (DSR) based multithreaded Semantic Focused Crawler (SFC) was proposed and presented in chapter 4. This crawler was used to fetch semantically relevant web pages from the Web on a given topic. The SFC used DSR to prioritize the web pages for setting the path for further crawling. DSR was computed during the crawl for each web page, based on the expanded list of the topic and the semantic distances among various semantically linked concepts from the domain ontology. Domain ontology was constructed manually on a few learning subjects, to include most of the related concepts which were linked based on their semantic relations. The potentially relevant web pages found by the SFC were stored in a local database.

The two ranking approaches, one to rank the untagged web resources by analyzing their content, and the other to rank tagged web resources based on their tags and popularity in the community, were presented in (chapter 5). The ranking of
untagged resources, those crawled from the WWW through Semantic Focused Crawler was computed using the proposed Semantic Term Frequency (STF) based approach, called CSR (Content based Semantic Ranking). On the other hand, in order to rank tagged web resources, those gathered from the Social Bookmarking System by FCHC crawlers, an algorithm called Social Semantic Ranking (SSR) was proposed, based on the Social Semantic Relevance (S2R). Although both approaches were based on the Vector Space Model (VSM), they were different in the selection of their dimension parameters in the vector space.

A mentoring task force consisting of primarily six types software agents and their assistants was designed to: i) manage sub-processes of web information retrieval; ii) organize the crawled resources semantically and; iii) provide them to a mentee on his/her request (Chapter 6). These software agents were designed with a capacity to create their assistant agents to perform their subtasks and handle multiple mentees without the delay in response time. The communication among the cooperative agents in different scenarios was explained through sequence diagrams.

The experimental study (described in Chapter 6) was conducted in three phases. Each phase was composed of the implementation of the components and their evaluation/verification depending on the evaluation needs. The first phase focused on the construction and verification of the Concept Ontology. In the second phase the proposed crawlers were implemented and evaluated for their relevant content gathering efficiency. In the third phase web resources’ ranking which was recommended by the system was evaluated with the users’ judgment.

The Concept Ontology at the time of study consisted of 14 Classes, 21 Object properties, 4 Data properties and more than 50 individuals from Domain Ontology alone. There were 8 Sub Class axioms and 6 Sub object property axioms. The ontology had inverse, functional, inverse functional and transitive axioms on the object property other than the domain and range axioms. It was validated on OWL Ontology Validator (WonderWeb), which showed a successful validation for OWL DL and OWL Full.
Conclusion

Web resources collected by the Social Semantic Focused Crawlers were analyzed and compared with other Focused Crawlers. The Classic Focused Crawler which was taken as baseline focused crawler, did not use any semantic or social information during crawling, and Semantic Focused crawler used only semantic information during crawl. Both these crawlers, Classic Focused Crawler and Semantic Focused Crawler fetched resources (or URLs) from the WWW whereas the variants of Social Semantic Focused Crawler (FCHC-BFP, FCHC-DFP-L1 and FCHC-DFP-L2), were designed to collect resources from the SBS based on different searching patterns.

All crawlers under experimental study were evaluated on the computed Semantic Relevance of crawled web resources using the well known performance metric, harvest rate. In spite of a large number of crawled resources, FCHC-DFP-L2 resulted in harvest rate of 98.5% and comparatively much lesser number of crawled resources, FCHC-DFP-L1 gave 99.3%; FCHC-BFP showed 97.1%, all with relevance threshold greater that zero.

Overall, The Social Semantic Focused Crawler with all its three search patterns showed better performances over Semantic and Classic Focused Crawlers. It was apparent from the experimental study that crawling SBS for relevant resources yielded better results as compared to the WWW crawling. However the relevant web pages those had not been bookmarked or tagged by any user could not be approached through the FCHC crawler. Thus to overcome the limited crawl area, a multithreaded DSR based SFC was designed.

The SFC crawler was evaluated using seed URLs from a search engine on various topics. An open source multithreaded focused crawler, Crawler4j was customized to crawl web pages on FCFS basis, hence was named as Classic FC for experimental purpose. The results obtained from both crawlers were compared and evaluated for the same data set with the same evaluation metric. The SFC which used domain ontology and DSR for crawling web pages performed better (showed improvement of 55% on an average.) for the given topics over Classic FC crawler.
7.2 Future Directions

This research similar to most research unwrap several issues for the future work. Keeping in view the pace of changing technology and the needs of users, the system was designed to accommodate expansions with time. The system has enormous scope to the expansions as it comprises of the technologies such as the Semantic Web and Multi Agent Systems. The Semantic Web which forms the basis for an intelligent futuristic Web is an emerging field of research. It, if opens research avenues at one end, then MAS at the other end crafts that research with a possible convenient implementation. The important issues that could be considered for the future work are as follows.

Capturing Multimedia Information: There is a lot of scope in the web crawling to find educational resources in varied formats. The thesis being bounded to the limits dealt with the textual content only. However the Concept Ontology and the resulted system were designed to incorporate the contents with different formats.

Ranking the Retrieved Information: There is a need to rank multimedia content, once it is captured from the Web. Moreover, the content with multiple formats need a common metric to determine their relevance, irrespective of the associated parameters (such as social tags, popularity count etc.) to the resource content.

Informal e-Mentoring: The PAM (Personal Agent to Mentee) can be improved further by incorporating the adaptive features such as learning and personalization so as to make it behave intelligently even in the absence of the mentee.

Retrieval from the Semantic Web: The agents can be designed on the pattern similar to that of the system’s DbpediaAgent to read the semantic data from the Semantic Web. This would enable the Virtual Mentor to deliver more precise knowledge to the mentee.

Global consensus on Agents’ Communication Language: The agents working across the Semantic Web enabled applications need a global consensus on the communication language. This could be implemented in the form of message passing through APIs. Unlike the existing communication languages, the design of the required new language must support a pure distributed Multi Agent System without any centralized or main agent.