CHAPTER 8

CONCLUSION AND FUTURE ENHANCEMENTS

8.1 SUMMARY

This research has focused on developing a Trust-based Personalized Blog Ranking and Summarization (TPBRS) system, to aid e-learning. In this work, the technical blogs are collected from various blog sources, using parallel blog crawlers which take domain ontology as the input, and simultaneously collect blogs from blog sites. The common blog representation using BlogML improves the accessibility of blogs in a better way. The blogs are checked for their trustworthiness, using the global reputation score and the local trust value in the blog network. The trusted blogs are clustered using the HABC algorithm, which improves the purity of the cluster; it was also experimentally analyzed that ontology based clustering is a better way for grouping similar blogs. The system also proposes a new ranking algorithm which ranks the blogs based on the content. The BRA uses the fuzzy probability based CBI measure to rank the blogs. The proposed system retrieves the blogs relevant to the user’s interest, and ranks the blogs based on the quality of their content.

8.2 CONTRIBUTIONS OF THIS RESEARCH

- A domain specific ontology-based blog crawler has achieved better crawling results. A maximum of 50 blogs have been
collected per each subject defined in the input domain ontology, using a parallel crawler from various blog sources.

- All the collected blog contents were converted into a Blog Markup Language (BlogML) in order to present the blogs in a generic format. This BlogML includes all the tags required for associating the blog content. A XML schema was designed to validate the BlogML.

- A blog filtering system was designed to remove the non-English, redundant, incomplete and irrelevant blogs. A trust model has been invoked to compute the trust value for the blogs and to refine the collected blogs based on trust.

- A Hierarchical Agglomerative Blog Clustering (HABC) clustering technique was used to cluster the blogs based on the subject, and to organize them in the blog repository for effective retrieval.

- User Personalization has highly reduced the blog retrieval time as it only retrieves the blogs based on user profile and user search log.

- It is clearly observed from the precision and recall values of the retrieved blogs, that the Blog Ranking Algorithm (BRA) has given the top ranking results.

- The Blog summarizer of the TPBRS system performs the calculation of mean TF-IDF for various blogs on the same topic. It achieves the best summary by combining the highest mean valued sentences from different blogs.
A recommendation system based on the learner’s feedback has been implemented for assisting the learners. The system has been compared with the traditional learning system, and the evaluation shows that the blog oriented supportive learning system has attracted the learners.

8.3 JUSTIFICATION FOR THIS STUDY

The Trust-based Personalized Blog ranking and Summarization system applies user personalization that highly helps to achieve better blog contents needed for the users, based on his own interest. Mainly, the personalization reduces the blog retrieval time as it only focuses on very specific blogs for that particular user. In most of the earlier works, personalization has not been used to retrieve less number of blogs with more content relevance for the users. In this work, in addition to personalization, the personalized contents are also checked for their trustworthiness. Thus, it helps to evaluate the trust of the blog contents existing among the bloggers. Generally, most of the ranking algorithms have concentrated and worked, based on the link-analysis of the web documents. But, the Blog Ranking Algorithm (BRA) uses the similarity analyser, that checks the content relevance of blogs for the user-entered search query. The rank value (Content-Based Importance) of the blogs is calculated, and also the similarity between the blogs is measured and analysed using the cosine similarity measures. In common, except for a few techniques, most of the text or content summarizers make a summary based on simply merging the contents. The blog summarizer of TPBRS uses the measure for calculating the TF-IDF of sentences existing in the blog contents. Using the highest mean value of TF-IDF, a brief summary is created.
8.4 FUTURE ENHANCEMENTS

There are many features that can be further improved, in the work for different domain data. Experiments have been done to compare the qualities of search with standard blog search engines, and the results show significant improvement in TPBRS as it effectively personalizes, ranks and summarizes the blogs for the users.

In recent years, various e-learning standards and techniques have emerged, to liberalize the available information technologies as well as to empower the current growing learners. In addition, this work has contributed and marked out ways for creating effective semantic-based supportive e-learning systems for various domains and application needs. This work mainly focuses on aid e-learning, by improving the efficiency of the needed techniques.

There is still scope for utilizing these applied techniques at a college course level as well as for various domain needs. In future, this research work can be enhanced by considering ontology related blog search for different domains. Improved clustering techniques may be applied on the domain specific blogs. Multimedia-based e-learning solutions (vlog) may be enhanced using the approach proposed in this research work.