CHAPTER 8

CONCLUSION

The most challenging task in information retrieval is the way of storing, retrieving and displaying results to the user. The organization and retrieval become complex due to the enormous number of documents available online. In order to organize the information for fast retrieval with accuracy, a framework named SFAIR is proposed in this work. The SFAIR framework consists of the components, namely (1) Indexing, (2) Clustering, (3) Information retrieval and (4) Ranking. This thesis has five different phases, where the first four phases analyze the four different components and the last phase concentrates on the SFAIR framework. The phases and their descriptions are as follows:

- **Phase 1:** The first phase deals with the multilevel hashing technique for indexing the web documents and enough experiments are carried to show the complexity of the proposed multilevel hashing. From the results, it is observed that the proposed indexing technique has less memory usage and minimum time consumption.

- **Phase 2:** In the second phase, the documents are clustered based on the context of the phrases present in the document. Results presented in this phase show that the proposed clustering technique performs better than SVM-based clustering in terms of time and accuracy.
• **Phase 3:** The most important phase of this work is information retrieval method, dealt in Chapter 5. The method retrieves the document using the semantic nature of the documents through SPDA technique. SPDA also ranks the retrieved documents. Efficiency of the proposed technique is measured through experiments and the results of the experiments show that SPDA technique outperforms the SIM approach.

• **Phase 4:** The information is retrieved and ranked using the feature-based technique using the DPDC technique. Necessary results are provided to show the efficiency of the DPDC against Latent Semantic approach.

• **Phase 5:** This phase combines the aforementioned four methods and constructs a framework named SFAIR. This framework initially clusters the documents using the technique presented in Chapter 4. The clustered documents are indexed for fast retrieval using the multilevel hashing technique. The information is retrieved for a user query using feature-based retrieval method discussed in Chapter 6. The pertinent documents that are retrieved are ranked using the techniques presented in Chapter 5. Results in this phase describe the efficiency of the proposed framework. For experiment, the dataset that has geographic information is used.

### 8.1 SCOPE FOR FUTURE WORK

Though this work retrieves the best pertinent documents to the user for their queries, the following are the points to implement in future to enhance relevance of the proposed work.
1. In future this work can be extended to use the spatial relevance-based retrieval process along with textual similarity

2. This framework can be made adaptable to multilingual retrieval