CHAPTER 8
CONCLUSION AND SCOPE FOR FUTURE WORK

8.1 INTRODUCTION

In this chapter, a review of main results is provided and contribution of this thesis and scope for future work is identified.

8.2 CONCLUSION

In chapter 2, the need for context focused search is identified. The several findings from the existing work in the area were discussed. First the need for focused search is identified as a solution of Information Overkill. The various focused search techniques were divided into main categories: link based and context based. The existing approaches apply various mechanisms to identify the relevancy of web page in respect to query. But none of the technique has considered the context of the query words or words in the web pages based on polysemy property of words. Therefore, sometime results in irrelevant documents in response to a query, this has become the motivation of the work carried out in this thesis.

The work carried out in this thesis has considered the polysemy property of words i.e. the various contextual senses of words has been used as a base for context focused search. The objective of work carried out is to explore contextual senses of words present in query as well as in web documents to get the exact context of user query and that of web documents. Further, to enhance the search results by including selective back-link pages. To fulfill all objectives the architecture for Context Based Focused Search (CBFSE) engine has been designed and a prototype for the same has been implemented. The various issues and contributions discussed below:

1. To extract the various contextual senses of the words polysemy in nature. So that the documents can be evaluated in respect to different contextual senses and search can be carried out in specific context only. To handle with this issue following contributions were made:
   - The WordNet dictionary has been integrated to get the various contextual senses.
   - A component named Contextual Sense Extractor (CSE) has been added in query processor module in proposed CBFSE.
2. To devise a mechanism where contextual senses of words can be used to get the context of the web documents and to assign a score that can be used to rank them. To address the same following contributions were made in this thesis:

- A context based relevance calculator mechanism has been proposed and implemented that compute probability of occurrence of a contextual sense in the web documents and assign a context score accordingly.
- A novel context based ranking of web documents is done based on the context score. The results are compared with Page Rank results.

3. It has been concluded from the existing literature available on back-links that back-links are the potential source of information in the same topical area. The objective is to check whether all the back-links are equally important to get the information in the same area. Thereafter, how selective back-links can be used to improve the search results. To incorporate the back-links in the results following contributions were made in this thesis:

- A mechanism to extract the back-links of downloaded pages is proposed and implemented
- The relevance of back-links w.r.t the associated web page is also computed and each back-link is assigned a relevance score.
- The effect of back-links over the ranking results is analysed only if the relevant back-links are included in the list of final results.

4. To address how the contextual information of words in the web documents can be stored in the index structure following contribution was made in this work:

- A context based index structure is proposed that stores the contextual senses associated with the words at a single place. It provides the provision to store the list of pointers to the related documents.

5. To get the specific context of the user query and then to display the relevant documents in that specific context of user query, this thesis has made the following contribution:
• A user interface is designed that displays the various contextual senses of the query keywords and allows the user to choose the specific context of his interest.

8.3 FUTURE WORK

Research work presented in this thesis can be extended along the following areas:

1. The context based focused search engine has used WordNet dictionary to extract the various contextual senses of words. The work can be carried out in future with the help of some other dictionary that provide functionality to extract various CSs.

2. The context based relevance evaluation mechanism can be applied to expand the query. The contextual senses integrated with some other features can be applied to query words to increase the query length in the selected context to improve the search.

3. The proposed context based mechanism is applied to handle only the textual data, the work can be carried out to include the functionality to handle video and images data.

4. The technique can be applied at crawler level to get the context of the web documents before the relevance evaluation is applied at query processing end to rank the web documents.

5. The context information can be augmented in the web documents and then that information can be used by any search engine while downloading the documents.

6. The proposed context based mechanism if integrated with other existing focused techniques will help the search engine to display more relevant as well as popular web documents to the user at top position in the result list. Thus, it can improve the precision to some extent.

7. The context based relevance evaluation of web pages can be applied in the distributed crawlers by just creating the instance of WordNet dictionary at each crawl agent. Thus, will help to download the relevant documents in the specific context of query.