CHAPTER 8

CONCLUSIONS AND FURTHER RESEARCH

8.1 SUMMARY OF CONTRIBUTIONS

This thesis reports an investigation into what additional information about a web page is required to be maintained in the index of search engines in order to improve the relevance of search results. The research has resulted in the design of a semantic search process, which aims to include the semantic information in the search process itself. Secondly the thesis also examines the crawler based updating procedure of the search engines and identifies the shortcomings of the procedure. The exploration ended in an autonomic computing based updating scheme, which aims to operate considering the internal and external factors of resource utilization. The summary of contributions include:

- Application of autonomic computing in search engine updating procedure
- Interpolation of delta transmission of updated web document
- Introduction of map model to depict the optimal path taken during updation
- Extension of search engine index to include semantic information of the web document.
- Formulation of an approach to append values to semantic index
8.1.1 Application of autonomic computing to search engine updation

Autonomic capabilities in a system accomplish their functions by taking an appropriate action based on one or more situations they sense in the environments. Web environment is dynamic; demands network resources to provide quality output for all web applications. Search engine, which is considered to be one of the popular web applications, has to produce results with the limited bandwidth available with them and with the web servers present in the global database. When search engines are imposed with autonomic capabilities, they try to adapt to the external as well as internal environments. The thesis focuses to provide autonomic capabilities to both the search engine server and the web server of the global database in the process of updating the repository.

8.1.2 Interpolation of delta transmission of updated documents

Delta encoding for HTTP (rfc 3229) states that when a web resource changes, the difference between the new and old instances can be sent to the client. Specification of delta encoding (rfc3229) does not include request messages or responses to methods other than GET. This thesis incorporates push technique instead of pull technique used by web crawlers to update the search engine repository. PATCH method is used to transmit the delta from the web server to the search engine server.

8.1.3 Appending semantic information to search engine index

Search engines generally maintain an inverted index of the web pages. The inverted index is the source to the search process to produce relevant search
results. The search process is mainly dependent on the index to produce the results. Context and category information of the web page is added to the index to narrow down the search to produce more relevant documents of users interest.

8.1.4 Approach to provide semantic values
Incorporating semantic values in the search process involves providing the semantic information at various stages such as in the index building process, in the index of search engines and in the information retrieval process. Entity headers to provide semantic information is introduced in order to build a semantic search index. The headers provide the semantic information in case of push or pull techniques of search engine repository building process. During the information retrieval process the semantic information is received as input from the user or it is derived from the previous browsing pattern of the users.

8.2 CONCLUSIONS
The concepts proposed in the thesis are tested for various parameters and the performance of the system is demonstrated using graphs. The performance is compared with the existing technologies and the benefits and limitations are recorded. The conclusions arrived by the analysis are summarized in this section.

- Traffic Adaptive Optimum updating Scheme (TAOS) records effective and efficient utilization of resources. The results show that the bandwidth
utilized in updating search engine repository using partial upload of the
updated document has reduced more than half of which is used during the
crawler based updation. The value of reduction in bandwidth utilization
ddiffers with the degree of change. Degree of change is directly proportional
to the bandwidth occupied during transfer, and hence an increase in delta
size will decrease the amount of bandwidth saved.

- TAOS improves the resource utilization by eliminating the needless
requests present in the crawler-based techniques. The results depict that
the bandwidth consumed by failed searches are more than consumed by
the successful searches. Notifying the updates through TAOS saves the
bandwidth consumed by failed searches. The autonomic computing design
helps in adapting to the dynamically changing environments, identifying
the best-suited page for notification.

- The precision of results demonstrated in the test results of semantic
search process shows improvement in user satisfaction. While there is no
difference between the scores provided by novice and professional users
in useful precision, best precision records variations in the scores. The
increasing document cutoff values recorded a comparatively less
deteriorating trend in the precision as that compared in the normal search
process.
8.3 LIMITATIONS

The current limitations of the research are as follows:

1. Limitation imposed by web servers - Web servers are suggested to have additional features to provide autonomic updation and semantic search process. These features are not present in the registered web servers and they have to be incorporated by the companies in order to find effective implementation of the proposed system.

2. Limitation imposed by protocol - Additional entity headers are proposed in the semantic search process and they are not present in HTTP. The importance of those headers in web information retrieval should be recognized and proposed entity headers should be included in HTTP to support effective implementation of the system.

8.4 RECOMMENDATIONS FOR FURTHER RESEARCH

This research work can be further extended in the following directions: autonomic computing based search engine repository building, universal information provider, cognition of user query, and extension on usability and scalability.

**Autonomic computing based search engine repository building**

Coverage of the web by the search engines can be increased by using an autonomic computing based top down approach. Autonomic computing should be implemented at various levels of transmission of the web pages including the network level. This may help to reduce the load imposed by the web crawlers in building the search engine repository.
Universal information provider

Universal information provider can be a repository of web pages which can be used by all search engines to build their index. This can reduce the network and server load created by web crawlers of different search engines. Different search engines can build their own index and apply their search and ranking algorithms to compete in the market.

Cognition of user query

User query should be refined before providing the input to the search process. An ontology-based categorization can be used to generate the user profile for the semantic search process. Cognition of user query will help to improve the search results by search engines.

Extension on usability and scalability

The research work can be further extended to a large scale with real users. This calls for conducting a usability and scalability study of the autonomic clients and servers of TAOS and semantic search process in a real life environment.