CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS OF WORK

This chapter summarizes the work of this thesis and lists its contributions. It also comments on possible directions of future work.

6.1 Conclusion
In this thesis we have proposed a multilayered framework for ontology driven web mining. The multilayered architecture is a three layered framework that implements clustering algorithms with WorldNet pre-processing.

We have analyzed clustering method with and without pre-processing step. The results depict that pre-processing with WorldNet ontology improves the results. The implementation is done using R tool and the metric used for evaluation is silhouette coefficient.
Following is a summary of the contributions of this thesis:

6.1.1 A multilayer architecture to web mining has been proposed. The three layered framework supports semantic clustering of document corpus by including WordNet ontology in the pre-processing phase of clustering.

6.1.2 The role of semantics and ontology for effective web content mining and information retrieval has been studied. It has been found that inclusion of semantics via ontology results in word sense disambiguation and improves results.

6.1.3 Implementation of various clustering methods on document corpus has been studied. The implementation has been done in R language and Python script has been used for WordNet pre-processing.

6.1.4 A comparative analysis of clustering methods with and without ontology has been done to evaluate the result of clustering method with and without WordNet ontology.
6.2 Future Directions

The possible directions of future work may aim to add further features in the proposed framework or improve proposed mechanisms. A lot of future scope has been observed during the implementation of the proposed framework.

The current implementation has been done on “The complete works of William Shakespeare” It can be enhanced on a larger corpus such as Wikipedia. The results can be compared with and without WordNet by applying various clustering methods.

Another future work may be to develop Twitter Ontology that can be used for opinion mining and sentiment analysis of twitter micro-blogs. Micro-blogging is one of the most popular Web 2.0 applications. Services like Twitter have evolved into a practical means for sharing opinions on almost all aspects of everyday life. The service shares and transmits huge volume of data. The task of extracting relevant and valuable information from the data is a challenge.

Micro-blogging web sites are rich data sources for opinion mining and sentiment analysis. Tweets typically do not consist of representative and syntactically consistent words. Text-based sentiment classifiers often prove inefficient. An ontology based framework can improve the results of opinion mining and sentiment analysis. The proposed framework will result in more detailed analysis of post opinions regarding a specific topic.