CHAPTER 6
CONCLUSION

Modern IR applications are barely depending the semantic similarity to retrieve the appropriate content from the vast information sources. WordNet is a widely recognized domain-independent lexical source of the semantic similarity measure. The value of the Wordnet is severely hampered due to the factors such as limited coverage, bias, and inconsistency. The quality of semantic relationships in WordNet is not appropriately updated because of the new practical usage and the new semantic relationships of words that have evolved over the period are left unnoticed. Thus, a concrete proposal for the enhancement of semantic similarity of words becomes the need of the hour which will improve the accuracy of IR.

The major contribution of the research is to develop a novel approach to enrich semantically the WordNet with the new relationship. The proposed approach semantically derives the evolving relationship from Wikipedia and web search engine aligned with the existing relations in WordNet. It executes a disambiguation mapping between each WordNet entry into its corresponding Wikipedia entry to retrieve a newly evolved relationship from the Wikipedia. In addition, a mapping algorithm is proposed to determine the number of additional senses available in Wikipedia pages rather than in WordNet alone. The proposed system also explains how it exploits the page count and the snippet returns by the Google search engine to disambiguate the word pair. It presents a lexical relationship retrieval algorithm and sequential pattern grouping algorithm to extract and arrange the number of lexical relationships the text snippets retrieved from a search engine. A two-class SMO-SVM is trained to select the suitable pattern that extends a new relationship to the ambiguous word in WordNet.

The performance of the proposed system reveals the effectiveness and accuracy of the approach. Experimental results outperform existing semantic similarity measure and achieve a correlation value of 0.87% close to human ratings in the benchmark dataset. Moreover, the real world applications such as IR and
query suggestion also demonstrated that the extended WordNet improved the accuracy of the results. The simulation results show that the proposed system enhances the correctness of the results in terms of precision by 5% as compared with the existing system. Thus, it is clear that the enriched WordNet provides measurable benefits in linguistic analysis, NLP and IR.