CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

The emerging trends in software development pose new research challenges in RE. Poor RE techniques are regarded as the biggest reason for software project failures. RE is an inherently complicated, multidisciplinary and multidimensional discipline which requires specific solutions by combining knowledge from different disciplines. The existing methods for requirements specification are not completely satisfactory. Hence this research investigated the advances in RE and its relation with different technologies, explored the domain knowledge and applied semantic knowledge to different classes of RE problems and finally enhanced the requirements written in natural language.

In this thesis, an integrated approach for the enhancement of natural language requirements which structured the domain knowledge by identifying key concepts and their relationships was proposed and implemented. The approach also included an in-built semantic assistance to utilize the domain knowledge and thereby enhanced unambiguousness, abstraction and clustering of requirements. Here, the theories and technologies of RE, semantic and domain analysis were combined and applied to the cases, LCS and MSS.
Initially, the requirements document was subjected to nocuous ambiguity detection and semantic similarity detection. The ambiguities were automatically analysed for the given natural language requirements document. A list of antecedent candidates were extracted from each anaphoric ambiguity instance. Possible coreference relationships among the NP candidates are identified using an NP coreference resolution engine. The coreference classifier used naive bayes algorithm algorithm to identify the NP combination in the dictionary and compared it with the OpenNLP models. The inclusion of semantic based heuristics improved the precision of coreference resolution by 3%. The antecedent classifier used the corefering NPs to classify the pronouns. The nocuous ambiguity judgement module used the naive bayes based antecedent classifier to distinguish nocuous anaphoric ambiguity from the innocuous ambiguities. The usage of brown corpora and wordnet dictionary improved the accuracy of the system and the system produced accurate results when the threshold is between 6.5 and 8.

Semantically similar sentences were identified with syntactic matching semantic similarity algorithm. Quality evaluator module evaluated the quality of the requirements document by measuring the quality indicators. The results showed that the proposed approach effectively detected the ambiguities, semantically similar sentences and improved the readability and understandability of the document by 5% and 14% respectively.

Then, the requirements were enhanced to build a reference document which would help the requirements analyst to create an SRS. This reference document would contain the significant terms in the domain. These significant terms are extracted based on their relevance in the domain documents. Multiple documents are available in domain and semantically similar terms may be portrayed in different forms in different documents. Hence, Semantic pheromone swarm algorithm was proposed and implemented
which tracked the terms across different documents. This was used for candidate link generation for similar terms and hence the extraction of single and multiwords. Experiments showed that the proposed method produced better Precision and Recall values with 2% improvement in the precision for the extraction of terms. Also, the exploitation of domain knowledge had enhanced the performance. When the number of terms is greater than 76, relevance based abstraction yielded better results. Hence the proposed method is well suited to large requirements documents.

Finally, the requirements were grouped based on their semantic dependencies and user priorities. The preferences of the requirements were collected from the user and then processed. The requirements priority matrix was constructed. AHP was applied to rank the requirements. A goal model was constructed and the semantic dependencies between the requirements were analysed and the requirements are clustered by k means algorithm. The requirements of every cluster were sub clustered based on the rank of the requirements. The performance of the clustering algorithm was enhanced by incorporating sampling techniques into it. Proportional stratified sampling had resulted in minimal variance clusters. Experiments and results showed that the clustered requirements matched the services very well with 4% improvement in the accuracy and therefore enabled efficient reuse of requirements and services.

Thus a semantic based integrated approach for the enhancement of natural language requirements was proposed and implemented. The approach explored the domain knowledge and used semantic based techniques for producing clusters of prioritized unambiguous requirements. The quality factors were evaluated in the initial stage and provided a scope for measuring and improving quality. Thus terms and their relationships were extracted from unambiguous documents with higher Precision and better Recall values thereby addressing the major challenges of the evolving RE discipline.
6.2 FUTURE WORK

Though the present research addressed the major RE challenges such as ambiguity, prioritization, domain dependence, reuse etc., the following enhancements can also be considered to yield a better RE framework.

- Latent semantic analysis could be applied to the requirements document.
- Modelling and capturing of non functional requirements such as security and privacy could be incorporated.
- Ontologies could be used to model the semantic dependencies between the requirements.