ABSTRACT

The emerging trends in Software development pose new research challenges in Requirements Engineering (RE). As the Software Engineering field evolves with service orientation, the applications are becoming more distributed and domain specific. Re-conceptualizations, adaptations and extensions are needed to the conventional RE techniques. Hence there should be a rethinking in the field of RE.

RE is an inherently complicated, multidisciplinary and multidimensional discipline which requires specific solutions by combining knowledge from different disciplines. The final output of RE is the Software Requirements Specification (SRS), which is the document that clearly and precisely describes each of the essential requirements of the software. SRSs are usually written in natural language which is highly ambiguous. The other RE challenges are domain dependence, prioritization, reuse and quality evaluation. The main issue in the existing RE techniques is that they do not handle the semantic aspects of requirements.

In this work, an integrated RE approach which enhances the natural language requirements by detecting the ambiguities and identifying key concepts and their relationships is proposed and implemented. The approach also includes an in-built semantic assistance to utilize the domain knowledge thereby enhancing unambiguousness, abstraction and clustering of
requirements. Here, the theories and technologies of RE, semantic and domain analysis are combined and used.

Initially, a Natural Language Processing (NLP) based ambiguity detection system is constructed that automatically analyses the ambiguities and enhance the given natural language requirements document. Syntactic matching semantic tree algorithm is proposed for the analysis of semantic similarity of sentences. Quality indicators along with quality measures find the defects and provide suggestions for improving the quality of the requirements. The results proved that the inclusion of semantic based heuristics improved the precision of coreference resolution by 3%. The usage of brown corpora and wordnet dictionary improved the accuracy of ambiguity detection and the system produced accurate results when the threshold is between 6.5 and 8. The readability and understandability of the document are improved by 5% and 14% respectively after the elimination of ambiguities.

Secondly, an extraction technique is proposed that effectively retrieves the key terms significant in a domain and creates a reference document which helps the analyst to acquire knowledge about the domain. The relevant terms in a domain may be either single or multiword. The extraction of the multiword is a difficult process due to the structural ambiguity of English. The semantic relatedness between the terms is also a problem in the extraction of terms. To overcome this, semantic pheromone swarm algorithm is proposed. This algorithm extracts semantically relevant terms and generates a graph that links the similar terms. Relevance based
abstraction is employed in which terms are extracted based on their occurrence in the document. The experiments showed that the significant terms and their relationships are extracted with higher Precision and better Recall values. The Precision of term extraction is improved by 2%. Relevance based abstraction provided better results when the number of terms is greater than 76.

Finally, a goal model consisting of high level goals and concrete level goals is constructed from the requirements document. Semantic dependency analysis is done to model interdependencies of goals. Then the goals are clustered, based on the usage, nonfunctional and threat dependencies. In every cluster, priorities of the requirements are processed and Analytical Hierarchy Process (AHP) is applied to rank the requirements. The ranked requirements are again clustered with sampling based clustering algorithm. These clusters can be used for software reuse, services selection and composition. Thus, the requirements are clustered based on their priorities and the dependencies between them. Priority based clustering results in 4% improvement in the matching index of the clusters.

Thus a semantic based integrated approach for requirements enhancement and knowledge extraction is proposed. This approach is illustrated and validated on two different case studies. Semantic based techniques are applied in all the stages. Key concepts and their relationships are extracted from high quality unambiguous requirements documents which serve as candidates to model the system.