1.1 REQUIREMENTS ENGINEERING

Requirements Engineering (RE) is a significant activity in the software engineering process. It is the process of identifying system stakeholders and their needs, defining constraints on the software system and documenting these in a form that is suitable for analysis, communication and subsequent implementation of the system (Roger 2005). The elicitation, analysis, negotiation, specification, validation and management of requirements are fundamental for the development of high quality complex software, irrespective of the nature of the software (Qadir et al. 2009). By fully understanding the stakeholders’ needs and documenting them in a concise and unambiguous way, it is possible to consistently deliver quality products designed to meet the complexities of the advanced information society.

A requirement is defined as a condition or capability needed by a user, to solve a problem or achieve an objective or that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed documents. This definition reflects the nature of RE as a multidimensional discipline as shown in Figure 1.1 (Berry et al. 2005).
RE has an inherently complicated, multidisciplinary and multidimensional nature which has made it a unique discipline requiring specific solutions by combining knowledge from different disciplines. RE is not only related to technical issues but also to managerial, organizational, economic and social issues. Accordingly, RE may not only be a front end process, but also be a part of the later stages of software engineering. Increased awareness about the importance of RE had resulted in significant RE research over the last 15 years. However, RE faces the following challenges (Insfran et al 2002).

- There is not one problem in RE, but a network of related problems. Thus, the problem boundary is vague and has to be identified during the RE process.
- RE intersects with many disciplines, i.e. it involves social, cultural, economic, environmental, organizational, political and even psychological issues.

- Many aspects involved in the RE process interact on various levels. For example, the activities, the people and techniques in the RE process interact with each other. Additionally, different business concerns and constraints of the software project interact with each other.

- Different application domains and different characteristics of specific projects make it impossible to apply the same technique to all different cases.

Using appropriate RE process models and techniques in a project is the first step towards increasing the overall quality of a software product. There is a close relationship between customer satisfaction and software quality. This is shown in Figure 1.2 (Pyzdek 2000).

![Figure 1.2 Relationship Between Customer Satisfaction and Quality (Pyzdek 2000)]
The expected quality can be obtained only when the customers are satisfied and the requirements are met.

The final output of RE is Software Requirements Specification (SRS), which is the document that clearly and precisely describes each of the essential requirements (functions, performance, design constraints and quality attributes) of the software and external interfaces. A good software requirements specification should be unambiguous, complete, verifiable, consistent, modifiable, traceable and usable.

In order to ensure the quality of a SRS, there needs to be a strong emphasis on implementing engineering disciplines into the RE process. It can be done by using various good practices, techniques and methodologies. But there are several reasons which hinder practitioners to extensively adopt RE. They are,

**Complexity of the problem domain**

RE techniques have to be adapted to the highly complex problem domains in order to elicit, model, document, verify and validate requirements. This is because no single method or technique is adequate for all situations.

**Unawareness of the existing RE techniques**

Research has shown that many practitioners are unaware of the range of RE methods available.

**Numerous and different levels of maturity of RE techniques**

Some RE techniques are relatively mature and well defined and have been used in various applications. Others are less mature and are rarely used. Due to the lack of guidance in choosing the RE techniques, it is very difficult for RE practitioners to apply RE techniques correctly.
1.2 ROLE OF RE

RE plays a crucial role throughout the entire software engineering lifecycle. A lot of research has shown that failures of most software projects are related to poor requirements. Similarly, it is not possible to develop better quality requirements without a well defined RE process. Based on empirical investigations and industrial experiences, numerous researchers emphasized that RE process is an essential contributor to the overall quality of the software product.

According to chaos report (The Standish Group 1994), good RE practices contribute more than 42% towards the overall success of a project. Improper RE practices account for over 43% of the reasons for late or over budget projects. The Chaos report analyzed more than 8,000 projects and estimated the total failure costs. The study conducted in 2009 in United States revealed that nearly two third of all Information Technology (IT) projects fail or run into trouble as shown in Figure 1.3 (The Standish Group 2009).

![Figure 1.3 The Standish Group 2009 – Chaos Report](image-url)
In addition to actual failure descriptions, the study also suggested project success, challenged and impaired factors as shown in Table 1.1. The table clearly shows the importance of complete and well-defined requirements in software development.

**Table 1.1 Project Success, Challenged and Impaired Factors**

<table>
<thead>
<tr>
<th>S.No</th>
<th>Success Factors</th>
<th>Challenged factors</th>
<th>Impaired Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>User involvement</td>
<td>Lack of user input</td>
<td>Incomplete requirements</td>
</tr>
<tr>
<td>2.</td>
<td>Executive management support</td>
<td>Incomplete requirements and specifications</td>
<td>Lack of user involvement</td>
</tr>
<tr>
<td>3.</td>
<td>Clear Statement of requirements</td>
<td>Changing requirements and specifications</td>
<td>Lack of resources</td>
</tr>
</tbody>
</table>

A case study carried out by Hall et al in 12 companies at different levels showed that out of a total of 268 development problems cited, 50% were requirements problems (Hall 2002). On the basis of studies conducted in hundreds of organizations around the world, it has been found that the RE practices are deficient in more than 75% of all enterprises (Jones 2000). Additionally, there are studies reporting the misunderstanding of requirements as the third most important risk factor for software projects (Keil et al 1998).

Incomplete requirements have been regarded as the biggest reason for software project failures. Miscommunication between business and IT contributes to 66% project failure rate. 34% of the IT Budget is wasted due to poor requirements definition and management (Information Architecture Group Consulting 2013). 60% to 80% of application or project failures traced back to poor requirements (Sjoberg et al 2007). 25% to 40% of all spendings on projects is wasted as a result of rework (Carnegie Mellon University-Software Engineering Institute 2013). All these studies suggest that RE plays a key role in software development.
1.3 RE PROCESS MODEL

A spiral RE process model was proposed by Sommerville and Kotonya (1998), which is shown in Figure 1.4.

![Spiral RE Process Model](image)

Figure 1.4 Spiral RE Process Model (Sommerville and Kotonya1998)

The Spiral model includes the following activities.

**Requirements elicitation**-Requirements are collected from different stakeholders.

**Requirements analysis and negotiation**-Requirements are analyzed and the problems identified are negotiated with the stakeholders.

**Requirements documentation**-Requirements are documented in appropriate notations.

**Requirements verification and validation**-Requirements are verified and validated to ensure that the requirements are real needs of stakeholders and are well defined.
**Requirements management**—Requirements management is highly related to requirements change management and includes requirements traceability.

1.3.1 **Requirements Elicitation**

Requirements elicitation is the process of identifying needs and bridging the disparities among the involved communities for the purpose of defining and distilling requirements to meet the constraints of these communities (Ryndina 2005). The other names for requirements elicitation are requirements acquisition, requirements capture, requirements discovery, requirements gathering, problem analysis and understanding, etc. This process requires application domain knowledge, organizational knowledge, technical knowledge and specific problem knowledge.

1.3.2 **Requirements Analysis and Negotiation**

Requirements analysis and negotiation is a process during which requirements are analyzed and modeled. The possible conflicts between the stakeholders are resolved by negotiations. The elicitation process provides the input to this process. The output of analysis is a consistent and complete set of requirements. Requirements can be modeled using three different kinds of languages or notations:

**Informal language**—Natural language is the most widely used informal modeling technique. It has an intuitive syntax and semantics, and can be easily understood and used. One of the major problems of requirements modeling with an informal language is that it can easily cause ambiguity and inconsistency of requirements.

**Semi-formal language**—Semi-formal language has a formal syntax but informal semantics. Typical examples of semi-formal language are
Conceptual Modeling Language (CML) (Brodie 1984) and Unified Modelling Language (UML) (Dennis 2002).

**Formal language**-This category of language has a well-defined formal syntax and semantics. It is built on sound mathematical theory, such as set, logic or algebra theory, or a mixture of them. A typical example is PetriNet. One of the major advantages of requirements modeling with a formal language is that it enables automatic analysis and verification of the requirements.

SRSs are usually written in natural language, often augmented or enhanced by information in other notations, such as formulae and diagrams. Natural language is preferred to write every initial conceptual document and every request to proposal. An online survey of the businesses requiring software, conducted at University of Trento in Italy showed that a majority of documents available for requirements analysis are provided by the user or are obtained by interviews (Daniel et al 2003). Moreover,

- 71.8% of these documents are written in common natural language,
- 15.9% of these documents are written in structured natural language, and
- Only 5.3% of these documents are written in formalized language.

### 1.3.3 Requirements Documentation

Requirements documentation is the process of documenting the agreed requirements at an appropriate level of detail in the most suitable notation based on a well defined document structure. This process has a very close relationship with requirements management.
1.3.4 Requirements Verification and Validation

Requirements verification and validation is the process of examining the requirements document to ensure that it is unambiguous, consistent and complete and that the stakeholders are satisfied with the final requirements specification.

1.3.5 Requirements Management

Requirements management is the process of identifying, organizing, documenting and tracking changing requirements in a project as well as the impact of these changes. It is an ongoing task throughout the whole RE process and might span as much as the whole software lifecycle.

Thus various activities are involved in RE. This is depicted in Figure 1.5 (Damian et al 2003).

![Figure 1.5 RE Activities (Damian et al 2003)](image-url)
RE is considered to be a difficult task due to the following reasons.

- Requirements are usually abstract and not well-defined at the beginning of the analysis phase making it difficult to extract a correct, viable and complete requirements statement.

- Analysts need to gather requirements information from various stakeholders, which cause problems like conflicting requirements, integration and validation problems. The need to document these requirements in order to relate them to their source stakeholders is crucial for subsequent processing and validation.

- Requirements must be developed and influenced by the environment that will host the system (organization context), which requires clearly identifying the environment, constraints that are imposed by the environment domain and the complexity of the environment itself especially when physical aspects of the environment need to be taken into consideration.

- The evolving nature of the environment demands requirements to evolve. Therefore, maintaining integrity and consistency among the already set and the new requirements is crucial for the continuity of the project and its ability to conform to various changes in future.

- It is impossible to frame requirements statements to be understood by both stakeholders (nontechnical) and domain experts (technical) people. The use of adequate processes and modeling notations is important to deliver documents which are easily understood by high level users and detailed enough for the usage of engineers.
1.4 NEW RESEARCH CHALLENGES IN RE

The new RE research challenges arise from the emerging trends in software systems and predictions about future software needs. Current trends in software development include increased scale of software systems, improved integration between software and its environment, greater autonomy of software to adapt to its environment and increased globalization of software development (Perry et al 2001). These trends reflect the changes in stakeholders’ needs and directly affect RE processes and practices. In some cases, current technologies can accommodate the new trends in requirements. In other cases, the trends pose new research challenges in RE, or increase the priorities of longstanding research problems.

1.4.1 Distributed RE

Globalization poses two main challenges to the RE research community (Ishraq et al 2012). The first challenge is that, new or extended RE techniques are needed to support the outsourcing of development tasks, such as design, coding and testing. Distance aggravates the gap between the requirements and development teams, particularly if the teams are from different organizations, have different cultures, or have different work environments (Herbsleb & Moitra 2001). As the geographic distance reduces team communication, the improperly defined requirements are at the risk of ultimately being misinterpreted, resulting in a system that does not meet the stakeholders’ needs. The second challenge is to enable effective distributed RE. Future requirements activities will be globally distributed, since requirements analysts will likely be working with geographically distributed stakeholders and distributed development teams may work with in-house customers (Herbsleb & Mockus 2001).
So practitioners would require techniques to facilitate and manage distributed requirements elicitation, distributed modeling, distributed requirements negotiation and the management of distributed teams i.e. not just geographically distributed, but distributed in terms of time zone, culture and language (Kraut & Streeter 1995).

1.4.2 Service Oriented RE

Service Oriented Architecture (SOA) is a collection of services that communicate with each other delivering service or value to subscribers over the Internet. As more and more service oriented projects are invested in, there arises a need to clearly understand and identify how requirements of such services can be developed. Though a lot of research has been made in the area of service oriented models and techniques, the related RE activities were not adequately considered.

In a service oriented environment, the requirement activities had to consider business aspects before considering the functionality aspects of the system. Researchers outlined that requirements are affected by both service providers and customers. The providers will try to design services which can suit multiple customers and to standardize services that suit all the customers. At the same time, the customers are looking for services that accommodate their needs. Based on service providers, they will have to change their processes to suit these services and therefore RE must be capable of complying with all these issues. The research of service oriented RE is promising and can open the door for a lot of both solution based and evaluation based research (Lichtenstein et al 2004).

1.5 RE RESEARCH IN ACADEMIA AND INDUSTRIES

Many universities, research groups and industries are actively involved in RE research worldwide. Some of them are listed below.
1.5.1 Software Engineering Institute, Carnegie Mellon University

The Research and Development (R&D) of Carnegie Mellon University focuses on core areas of RE. It works on RE for agile development. The core areas of research include requirements management, systems management and system construction. The research aims to specify the ways in which the organization and project management environment for system development can support or reject improved quality requirements elicitation mechanisms (Carnegie Mellon University - Software Engineering Institute 2013).

1.5.2 RE Group

RE group at University of Technology, Sydney (RE@UTS) was established in 2001. The overall aim of RE@UTS is to increase the awareness of both academics and practitioners about the importance of RE in software and systems development and to foster high quality RE research and training (University of Technology 2013). The group is aimed at informing people of the range of research projects, training programs and consulting services in RE and also to provide access to a collection of worldwide RE resources.

1.5.3 RE Research Group

RE Research Group (REQENG) is constituted by the UNSW Australian School of Business and is aimed at examining the current practices used for RE in various industrial settings in terms of how it is done and what needs to be done to improve the quality of requirements specification, which leads to mutual benefits for implementers as well as end users (Australian School of Business 2013). The focus areas of research are,

- RE in global software development
- RE in agile development
- RE and lean development
- Decision making in RE
- Value based software development

1.5.4 The Irish Software Engineering Research Centre

The Irish Software Engineering Research Centre’s (LERO) research in RE covers the development of novel techniques for the elicitation and analysis of requirements and relates requirements to other processes and artifacts in the software development process (The Irish Software Engineering Research centre 2013). The areas marked for research by LERO are:

- Analysing natural language requirements
- Formal specification and analysis of requirements
- Managing inconsistency and evolution of requirements
- Relating requirements and design
- Managing quality requirements, including safety, security, privacy and usability requirements
- Modeling, analysing and enforcing requirements for adaptive systems.

1.5.5 The FARE Project

University of Wollongong is engaged with the FARE (Formal Aspects of RE) project. This project seeks to address the deficiencies in RE by developing a formal framework that includes a conceptual modelling language at the front end and an underlying formal knowledge representation language. The framework helps to define the normative requirements criteria
and processes such as evolution and management of inconsistency and multiple viewpoints (University of Wollongong 2013). This would serve as the basis for a RE toolkit which provides comprehensive automated support with the benefits of clear semantics and formal verifiability.

1.5.6 University of Zurich

Rigorous RE research is undertaken at University of Zurich and Table 1.2 lists some of the ongoing projects (University of Zurich 2013).

Table 1.2 RE Projects

<table>
<thead>
<tr>
<th>S.No</th>
<th>Title of the Project</th>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>ADORA</td>
<td>Development of a semiformal specification language and a prototype tool that overcome the weaknesses of UML.</td>
</tr>
<tr>
<td>2.</td>
<td>AquaBench</td>
<td>A benchmark for requirements traceability.</td>
</tr>
<tr>
<td>3.</td>
<td>Flexible Modeling</td>
<td>To unite the flexibility of unconstrained sketching with the power of formal modeling.</td>
</tr>
<tr>
<td>4.</td>
<td>MiRiA</td>
<td>To define objective measurement of a model's abstractness and systematic guidance to attain the right level of abstraction.</td>
</tr>
<tr>
<td>5.</td>
<td>Requirements update</td>
<td>To facilitate the update of requirements documents when software systems evolve.</td>
</tr>
</tbody>
</table>

1.5.7 Auckland University of Technology

Auckland University of Technology (AUT), New Zealand works on RE research projects related to the understanding of practitioners’ problems, improving processes, their adoption and providing tool support (Auckland University of Technology 2013). The research is related to RE
activities such as requirements elicitation, modelling, analysis, tracing, prioritization, verification and validation, reuse and change management. The AUT group is interested in a cognitive view of RE and cognitive support of RE activities.

### 1.5.8 Tata Consultancy Services

The research group at Tata Consultancy Services (TCS) is developing a new approach to RE which is termed as Knowledge Assisted Requirements Evolution (KARE) (TCS 2013). This approach focuses on several sub-problems of RE such as,

- Structuring domain knowledge by way of identifying key concepts, their relationships and constraints
- Presenting the knowledge in a way that makes it visible and easily accessible
- Providing an in-built semantic assistance to utilize this knowledge and thereby enhance unambiguosity, completeness and correctness of requirements specification
- Provide a way to create requirements models amenable to downstream development
- Providing effective collaboration mechanisms for highly interactive activities.

### 1.5.9 International Business Machines Corporation

Research at International Business Machines Corporation (IBM) is exploring both methodological and tool based approaches to address the challenges of multisite requirements management. A prototyped tool called
EGRET (Eclipse Based Global Requirements Engineering Tool) extends conventional requirements management tools with rich support for formal and informal collaboration, along with knowledge management capabilities (IBM 2013).

These research activities and projects highlight the potential of RE as a significant, challenging and crucial research area with many problems remaining unexplored.

1.6 MOTIVATION

Most of the researches in RE focus on new or improved techniques for evaluating the quality of recorded requirements. Some analyses look for well-formed errors in requirements, where an error can be uncertainty, inconsistency or incompleteness. Other analyses look for anomalies, such as unknown interactions among requirements, possible obstacles to requirements satisfaction, or missing assumptions. Both types of analyses reveal misunderstandings or questions about the requirements that usually call for further elicitation. Requirements analysis also includes techniques, such as risk analysis and impact analysis. These help the analyst to understand the requirements, their interrelationships and their potential consequences better. So, the analyst can make more informed decisions, visualizations and subsequently help a manager to select an optimal combination of requirements to be implemented, or to identify acceptable off-the-shelf solutions.

One of the major tasks of RE research is to help requirements engineers to develop a suitable RE process for a project so that quality requirements can be obtained (Asghar & Umar 2010). Multidisciplinary
approaches, which integrate software engineering, RE, knowledge engineering and Natural Language Processing (NLP) have to be used in order to provide feasible solutions to the problems identified (Ebert & Wieringa 2005).

As the software development life cycle progresses, if the requirements retrieved are not formally documented or analyzed or updated, the quality of the software will degrade. Requirements consistency checking, requirements tracing etc are done to improve the quality of the software. Therefore, to improve the quality of the software, researchers are involved in discovering better solutions.

Ambiguity is an essential feature to be considered as it affects the natural language requirements document and thereby affects the quality of the software. Many pre-processing activities are involved in carrying out the ambiguity detection and classification (Issa et al 2011). The natural language requirements documents are error-prone and large in size (Dag & Gervasi 2005). Before analyzing the requirements documents for the levels of ambiguity, the evaluation of requirements documents needs to be considered as it involves a significant role in analyzing the document characteristics.

However, the existing methods for RE do not satisfy all the criteria. Requirements must be measurable, testable, related to identified business and other field needs or opportunities and defined to a level of detail sufficient for system design. The demand for better, faster and more usable software systems will continue and RE will therefore continue to evolve in order to deal with different development scenarios. The objective of this work is to enhance the requirements written in natural language by investigating the advances in RE and its relation with different technologies, exploring the
domain knowledge and the application of semantic knowledge to different class of RE problems.

The semantic aspects of requirements are not considered by the existing methods (Breaux et al 2008). There had been intense research in the area of information extraction from a domain but these techniques have not focused on the use of domain expertise in RE. The difficulty in understanding knowledge based techniques is very high in RE since requirements documents are aimed at a range of stakeholders from different backgrounds and domains of knowledge. If automated techniques are used to consider the meaning of requirements, it is possible to obtain more effective RE techniques to produce high quality requirements specifications.

In this thesis, an integrated approach is attempted for the enhancement of natural language requirements. This will be used for structuring the domain knowledge by identifying key concepts and their relationships. The approach also includes an in-built semantic assistance to utilize the domain knowledge and thereby enhance unambiguosity, abstraction and clustering of requirements. Here, the theories and technologies of RE, semantic and domain analysis are combined and used.

In the first stage, ambiguous terms are detected and quality indices are evaluated. Then, domain knowledge is explored to generate the significant terms from a domain. In the third stage, dependency analysis is performed, the relationships are extracted and the requirements are prioritized and grouped. Semantic techniques are applied in all the three stages and finally key concepts and their relationships are extracted which serve as candidates to model the system.
1.7 CONTRIBUTIONS

The significant contributions of this research work are the following.

- A NLP based ambiguity detection system is designed to automatically analyse the ambiguities of a given natural language requirements document. Syntactic matching semantic tree algorithm is proposed for the analysis of semantic similarity of sentences. Quality indicators measure the quality index and provide suggestions for improving the quality of the requirements.

- A semantic pheromone swarm algorithm is proposed which effectively extracts the key terms significant in a domain and helps the analyst to acquire knowledge about the domain. Relevance based abstraction is applied and the semantically relevant single and multi words are extracted from all the domain documents.

- A goal model is constructed from the requirements document which depicts the goals and the semantic dependencies between the goals. A sampling based clustering method is proposed which groups the requirements based on user priorities and dependencies, thereby producing prioritized and semantically clustered requirements.

- The salient feature of this research is the exploitation of semantic knowledge in every stage of the enhancement of natural language requirements.
The overall system design is shown in Figure 1.6.
1.8 ORGANIZATION OF THE THESIS

The rest of the thesis is organized as follows and is elaborated in Figure 1.7. Chapter 2 discusses the state of the art of RE knowledge and summarizes the research in RE. The advancements in other allied fields related to the research such as NLP, knowledge extraction and semantic analysis are also discussed.

![Figure 1.7 Organization of the Thesis](image-url)
In Chapter 3, a method for producing quality requirements is proposed which includes an NLP based system for the detection of sentence end words, tokens, Parts Of Speech (POS), anaphors and coreferring phrases. The detection of anaphoric ambiguity and identification of the coreferring NPs are done automatically by the system. The system has the quality evaluation module to evaluate the quality of the natural language requirements document.

Chapter 4 proposes an approach for the extraction of key terms. NLP techniques are combined with relevance based abstraction. One of the main contributions of the approach is the semantic ranking of terms and use of pheromone swarm algorithm, which avoids the problems associated with semantic relatedness between the terms.

Chapter 5 proposes a technique for the clustering of semantically dependent requirements based on their priorities. A goal model is constructed from the requirements document and semantic dependency analysis is done to model and cluster interdependencies of goals. In every cluster, priorities of the requirements are processed and the requirements are ranked and clustered based on these priorities.

Chapter 6 presents the conclusion and future directions of the research.