CHAPTER-5

A NOVEL DESIGN APPROACH FOR WORD SENSE EXTRACTION BASED ON ITS CONTEXT

5.1 INTRODUCTION
In modern e-document technology, understanding the precise feeling (or significance) of the term in a record becomes an essential problem. It's considered that understanding of a real significance of the polysemous word may enhance the caliber of the data access procedure through locating more relevant files or removing relevant information in the files. Feeling multiple meanings in a big digital text is quite hard with a device as compared to organic human terminology. In normal language, the word sense is extracted by human by connecting it to that specific circumstance. But for digital text this function is performed by natural language running by removing two qualities of term.

i) It removes ambiguity of an individual word that can be used (in different contexts) to express multiple meanings.

ii) It identifies different meanings of word by extracting relation between two words that are spelled the same way.

Two assets are crucial: A Understanding for obtaining connection of word in framework and framework in that the word continues to be utilized, to feeling any word. Individual has an capability to discover connection understanding of term in a circumstance. For instance for a term "Fine" within the framework of individual situation it connect more word like "look", "nicely", "feel". Because of these related term human condition won't be definitely described by it rather than send to punishment. Because of insufficient understanding cleverness in pc, extra resources are used by it to feeling term like dictionaries, tagged files etc. WordNet is a broadly employed lexicon for the British vocabulary. It groups English phrases into sets of synonyms called synsets, supplies brief, basic meanings, and records the numerous semantic relationships between these synsets. In WordNet info is arranged in the conceptual significance level. As a synset which is then associated with additional ideas through associations each notion is known. Word feeling in WordNet is created by team of synset, where a synset of the term w is some w's synonyms. In WordNet
each term goes to 1 or more synsets as polysemic term Known. Following are principal methods (Roberto) utilized in pc to feeling term.

I. Dictionary-based Algorithms: It utilizes information assets in the shape of machine readable dictionaries to remove multiple feeling of term. Book identifies a phrase in a special area.
Advantages : unnecessary time is saved by It needed to produce training data.
Disadvantages : In this strategy all tagged sense doesn't be provided by it in one book, therefore method is required by it to include fresh sense in book.

II. Supervised Disambiguation Algorithms: It utilizes information source within the shape of labeled corpora which identifies meaning of term. It builds classifier which groups new term right based on their context useful. Large sense training is needed by it established to remove sense of term.
Advantages: when compared with additional strategies supervised algorithms provide the very best efficiency when it comes to correctness which may achieve quite nearer to sense removal.
Disadvantages: Supervised formula needs a substantial number of instruction information which requires additional time to produce it.

III. Unsupervised Disambiguation Algorithm: Unsupervised disambiguation algorithm is comparable to clustering where they group examples of the term by significance.
Advantages: Unsupervised algorithms uses training dataset for which marked sense opinion are not needed. Compared to supervised algorithm efficiency is higher.
Disadvantages: Compared to monitor algorithm poor accuracy is given by it.

5.1.1 Foundation of Related Technology
5.1.1.1 Mining Data:
Information exploration is the evaluation of substantial amounts of information, in order to recover significant and helpful patterns and principles. The amount of information is growing daily. To ensure that the correct and best effect is acquired, to be able to get the better of the inadequacies of guide evaluation, data exploration methods may be used. Information mining entails a chain of measures. In categorization, the incoming information is arranged by evaluating their characteristics for the predetermined aspects of a category. In approximation, a edge
limit is examined and proven whether the information worth is above or below that the categorization and limit is performed. Organization guidelines assists to determine which combos are greatest, so the result is better. In clustering the group of information is performed. There might be a number of comparable organizations. There's no predetermined information in clustering. The clustering is performed on the foundation of the commonalities in the information acquired. There's a broad variety of methods you can use to mine information. Mathematical methods, neural systems, machine learning methods, genetic algorithms, rough sets methods, fuzzy set methods, decision tree construction processes, k closest neighbor's methods, along with other resources can be found for information exploration. All these techniques has its flaws and advantages, and area of the worth supplied to the endeavor from the information exploration group lies in recognizing which techniques to make use of, and when.

5.1.1.2 Text Mining:

Text mining, also known as Knowledge-Discovery in Text (KDT), refers to the process of extracting interesting information and knowledge from unstructured text. [36] Data mining tools are designed to handle structured data from databases, while Text mining can handle unstructured or semi-structured data sets such as emails, full-text documents and HTML files etc. Fig. 5.1 depicts a generic process model for a text mining application.

Beginning with a group of files, a text exploration device might recover a pre and special file - process it by examining personality and structure models. Then it might undergo a text evaluation period, occasionally repeating techniques until info is removed. Two text evaluation and removal techniques are proven in the fig. 5.1, but several additional combos of techniques might be utilized based on the aims of the business. Like a design finding which is aid to translate objective knowledge the resultant info could be put.
5.1.1.3 Natural Language Processing:

Natural Language Processing (NLP) is a place of study and examines that program how computers may be used to comprehend and language natural control text or language to do helpful things. Normal language can be used to signify human ideas and human measures. Normal language processing produced systems that instruct computers to evaluate, comprehend, and actually create text. A few of the systems which have been created and may be used within the text exploration procedure are information extraction, issue tracking, summarization, categorization, clustering, theory linkage, information creation, and query addressing. Programs of NLP include numerous areas of studies, including cross language information access (CLIR), speech recognition, artificial intelligence etc. (Mansuy) and device translation, natural language text running and summarization, user interfaces, multilingual.

5.1.2 Software Quality Assurance:

“In the context of software engineering, software quality measures how well software is designed and how well the software conforms to that design.” Software quality guarantee is an umbrella exercise that is employed through the software procedure. Software quality can be defined as “Conformance to explicitly stated functional and performance requirements, explicitly documented development standards, and implicit characteristics that expected of all professional developed software”. Software Quality Assurance (SQA) is the method of guaranteeing folks that every attempt continues to be designed to make sure that software products
possess the desirable amount of dependability, maintainability, functionality, and scalability (Pressman).

Above explanation imply that applications quality assurance goals following three factors.

I. Software requirements will be the basis that quality is calculated. Absence of conformance to needs is lack of quality.

II. Given requirements determine some improvement requirements that direct the mode by which computer software is designed. Deficiency of quality may almost certainly result, when the requirements aren't adopted.

III. If software conforms to its explicit requirements but fails to meet implicit requirements, software quality is suspect.

5.1.3 Trends in Text Word Sense Extraction:

Looking back, therefore a lot of the first function on Word Sense Removal is documented in fairly not certainly indicated in posts and publications across professions and a few areas. With enough greater assets and increased mathematical techniques at their removal, scientists in the 1990's have clearly enhanced on earlier outcomes. Of program, Word Sense Removal is challenging in part as a result of the constitutional problem of identifying as well as defining word sense, which isn't a problem that's apt to be resolved soon. Among the apparent stumbling blocks in much current Term Sense Extraction function is the fairly thin perspective of sense that comes hand-in-hand with the effort to use sense variations in every day dictionaries, which can't, and aren't meant to, signify meaning in circumstance. Hopefully this will contribute to work on Word Sense Extraction. In common conditions, word sense removal in text includes the organization of the specified word in a text or discussion with a description or meaning (sense) which is distinguishable from additional connotations possibly due to that word (Ide). The job so always entails two steps: (i) the dedication of all the distinct feelings for every word applicable (at least) to the text or discussion under consideration; and (ii) a means to allocate each event of a term to the proper feeling. Much current function on word sense extraction depends on pre-defined senses for measure (i), including:

a) A checklist of senses including these found in regular dictionaries;
b) A team of attributes, classes, or related phrases (e.g., alternatives, as in a thesaurus);

c) An accessibility in a move book which contains translations in another language; and so forth.

5.1.4 Context Role

Circumstance may be the only real way to recognize the significance of the polysemous term. Consequently, all function on sense extraction depends on the circumstance of the mark term to supply advice to be utilized because of its sense extraction. In some instance, context also offers the preceding understanding with which present framework is compared to reach disambiguation. Generally speaking, circumstance can be used in two ways:

I. The bag of words approach

Here, context is considered as words in some window surrounding the target word, regarded as a group without consideration for their relationships to the target in terms of distance, grammatical relations, etc. Surrounding word of the target word decides correct choice of word sense.

II. Relational information

Context is considered in terms of some relation to the target, including distance from the target, syntactic relations, selection preferences, orthographic properties, phrasal collocation, semantic categories etc. Information from micro-context, topical context, and domain contributes to sense selection, but the relative role and importance of information from the different contexts and their inter-relations are not well understood. Overall, the bag of words approach has been shown to work better for nouns than for verbs, and to be in general less effective than methods which take other relations into consideration.
5.2 LITERATURE REVIEW

Plenty of function continues to be completed in these areas by scientists. This part of study enlightens briefly on some of work done by those researchers. The work from various books, papers, articles, journals has been referred for this purpose. It is humble approach to thank those researchers whose work done will be referred in this research. Some of them are mentioned in this article.

(1) Gupta and Lehal (2009)

The Writers current text exploration, also called Knowledge-Discovery in Text (KDT), describes the procedure of removing interesting info and understanding from unstructured text. While Text mining can manage unstructured or semi-structured data models such as emails, full-text records and HTML documents etc, information exploration resources are made to handle organized data from directories. Contextual meaning can be easily handled by human but pc can't handle easily spelling versions and contextual meaning of text till some guidelines supplied for the pc. When the connotations of a bit of text need to be removed in a special circumstance this situation becomes crucial and more essential. Natural language processing (NLP) can be used to find out which feeling of a term ought to be implemented for every single case of the term. Beginning with a group of files, a text exploration device might recover a pre and special file - process it by examining personality and structure models. Then it might undergo a text evaluation period, occasionally duplicating techniques until info is removed. Text evaluation and removal methods are proven in the figure 5.2. The resultant info could be put like a design finding which is aid to translate objective knowledge.

Figure 5.2: A text mining Generic process
The authors described about technologies used in text mining. At introductory part author tries to define and explain basics about text mining. The writers create about systems which have been created and may be used within the text exploration procedure like information extraction, issue tracking, summarization, categorization, clustering, theory linkage, and query addressing.

Information removal - In this area predefined sequences used to recognize associations and search phrases within text.

Topic Tracking - In topic tracking person account is kept and depending on the files the user opinions, forecasts other documents of interest for the person.

Summarization - Text summarization is exceptionally useful for attempting to determine whether a extended file fulfills the user's wants and it may be worth studying for additional advice.

Categorization - Categorization requires determining the chief styles of a record by putting the document into a pre-defined set of issues.

Clustering - Clustering is a method utilized to team related files.

Concept Linkage - related documents are connected by concept linkage tools by determining their generally-shared ideas.

Query Addressing - It handles how to locate the very best solution to some provided query.

Overall basic details required for text mining. It gives idea about preprocessing text files used in dissertation. It gives foundation for creation of important keyword by removing common stop words. It also gives basic details for mapping relation of keywords.

(2) Mocian (2009)

The author started with relation of Word Sense extraction and Natural Language Processing. Author describes that Word Sense extraction technique handles two properties of the English language, Polysemy and homonymy. These properties characterize a word which has multiple senses, and two different having the same spelling, respectively. WSE is employed whenever a semantic comprehension of text is required. Writer [40] also covers Phrase Sense extraction with regard to weaver (1949) who was formulates a really significant issue about the framework utilized for disambiguating words: How several contiguous words ought to be taken into consideration when disambiguating a particular word in text. As, so as to
disambiguate a term, two assets are crucial: some sort of understanding linked to the term, and a circumstance by which the term has been utilized. The writer describes it with instance. Individuals possess "world information," which aids them disambiguate words readily. For instance, when the term "bank" seems in a text that additionally includes phrases like "mortgage", "assets" or "bailout", we may understand that the word "bank" describes a lender rather than towards the property near a water. A term must be disambiguated only if it has several feelings. It gives information about word sense extraction in different contexts. It gives core idea for sensing word in different context by using its neighboring words. This idea is implemented by mapping neighboring words relation with WordNet dictionary synonyms.

(3) Ide and Veronis (1998)

The authors discuss overall survey and early work on word sense extraction in text. Author started with word sense discussion work by Weaver (1949) which describes the need for word sense extraction in machine translation. The very first efforts at automatic sense disambiguation were created in the framework of machine Weaver (1949) traces the foundation of a strategy to term sense removal in text the following. The useful issue is: “What minimum value of N (Neighboring word) will, at least in a tolerable fraction of cases, lead to the correct choice of meaning for the central word?” Then authors concentrates on early test with preceding query done by Kaplan (1950) which extracts sense of main term by getting a couple of phrases on both side of central term. The impressive truth about it early function on word sense removal is the amount to which the strategies and essential difficulties towards the difficulty were foreseen and created in those days. It gives overall earlier work on word sense extraction. It provides idea about selecting value of neighboring word for extracting word sense in particular context.

(4) Chen and Al-Mubaid (2006)

The authors try to say text mining technique for medical domain based on machine learning method. They say each meaning of a word is called a sense. In natural language processing (NLP) word sense disambiguation (WSD) is used to determine which sense of a word should be adopted for each instance of a word. Result is tested by considering Gene and protein as two different contexts. It gives text mining work in medical domain. Idea is useful for finding word sense by its each instance in context.

The authors focused on frequent word co-occurrence in single document to extract keywords. A document consists of sentences. A sentence is considered to be a set of words separated by a stop mark (‘.’, ‘?’ or ‘!’). The stop-words are high frequent words that carry no information (i.e. pronouns, prepositions, conjunctions etc.). Remove stop-words can improve clustering results. Author tested results for Similarity-based clustering and Pair wise clustering. It gives description about how to extract keywords. Idea is useful to improve performance by removing common stop words and putting only important keywords.

(6) Bajwa (2010)

The Author uses Markov Logic to extract context based meaning of word. Markov logic has ability to handle uncertainty and learn from the training data. Author says, various linguistic phases are common for processing natural language i.e. lexical and semantic analysis, pragmatic analysis Text parsing in NLP can be a detailed process. Some applications e.g. text summarization and text generation needs detailed text processing. In detailed process every part of each sentence is analyzed. The text mining formulates the standard speech language rules with certain weights. These meticulous weights for each rule ultimately support in deciding the particular meaning of a phrase and sentence. It gives detail about extracting word sense by means of noun, adverb, and adjective etc. Idea is useful for using extra resources as WordNet dictionary to sense word.

Author proposed architecture with three component modules.

a) Extracting valid NL terms and identifying their POS (parts of speech) type.

b) Markov logic based semantic analyzer has been used to understand meanings of the sentences.

c) Knowledge based system that consists of English word libraries; data type taxonomies and parse trees.

(7) Roberto (2009)

The author surveyed the field of word sense disambiguation. Human language is ambiguous, so that many words can be interpreted in multiple ways depending on the context in which they occur. Most of the time humans do not even think about the ambiguities of language, machines need to process unstructured textual information and transform them into data structures which must be analyzed in order to determine
the underlying meaning. The computational identification of meaning for words in context is called word sense disambiguation (WSD). Author distinguishes two main approaches to WSD:

i) Supervised WSD: These strategies use machine-learning methods to understand a classifier from tagged training sets, that's, models of examples secured in conditions of a quantity of attributes as well as their proper sense tag (or course).

ii) Unsupervised WSD: These procedures are centered on unlabeled corpora, and don't utilize any manually sense-tagged corpus to supply a sense alternative for a term in context. It gives detailed survey of word sense extraction with approaches.

It is useful to select approach for sensing word in context.

(8) Collobert and Weston (2008)

Author describes about NLP approach to evaluate human language in machine. The area of Natural Language Processing (NLP) aspires to transform human language into a official portrayal that's simple for computers to control. Information extraction, machine translation, summarization, search and human computer interfaces are included by current end applications. Scientists took a separate and conquer strategy and recognized several subtasks useful for evaluation and program development, while entire semantic comprehension continues to be a far - distant target. These are the syntactic, such as part-of-speech tagging, chunking and parsing, to the semantic, such as wordsense disambiguation, semantic-role labeling, named thing removal and anaphora resolution.

Author classified NLP Tasks into six standards NLP as follows.

Part-Of-Speech Tagging (POS) aims at marking each term having a distinctive label that signifies its syntactic function, e.g. plural noun, adverb.

Chunking, also called shallow parsing, aims at labeling sections of the word with syntactic components for example noun or verb phrase (NP or VP). Each term is designated just one distinctive tag, often secured as a start-chunk (e.g. B-NP) or interior-chunk label (e.g. INP).

Named Entity Recognition (NER) tags atomic components within the phrase into classes for example "INDIVIDUAL", "BUSINESS", or "PLACE".

Semantic Role Labeling (SRL) aims at providing a semantic role to a syntactic component of the phrase. E.g. the pursuing phrase may be labeled "[John] ARG0 [ate]REL [the apple]ARG1 ", where "ate" is the predicate. The exact arguments rely
on a verb's framework and when you can find several verbs in a phrase some phrases may have multiple labels.

Language Versions A language design typically estimates the chance of the following term being in a series. Writer views another setting: call whether the specified series exists in character, or maybe not, following a strategy. This really is accomplished by marking actual texts as favorable illustrations, and producing "imitation" negative text.

Semantically Related Phrases ("Synonyms") that may be the job of forecasting whether two words are semantically related (synonyms, hypernyms) which is calculated utilizing the WordNet database. It gives detail about evaluating human language on machine. It gives idea about mapping word relation with WordNet database.

(9) Reichart and Rappoport (2010)

The Authors explain word sense by means of Polysemy. Polysemy is a major characteristic of natural languages. Like words, syntactic forms can have several meanings. Understanding the correct meaning of a syntactic form is of great importance to many NLP applications. In this paper author address an important type of syntactic Polysemy the multiple possible senses of tense syntactic forms. Authors make their discussion concrete by introducing the task of Tense Sense Disambiguation (TSD): given a concrete tense syntactic form present in a sentence, select its appropriate sense among a set of possible senses. Author describes a supervised TSD algorithm trained on these annotations, which outperforms a strong baseline for the task. In this paper author address a prominent type of syntactic form Polysemy: the multiple possible senses that tense syntactic forms can have. Disambiguating the Polysemy of tense forms is of theoretical and practical importance. It explains word sense by expressing two or more different meanings of same word. It gives idea to interpret word sense according to its context.

(10) Manning and Schütze (1999)

The Writers describe about term sense disambiguation difficulty that have several connotations or feelings. For such phrases provided of circumstance, there was hence ambiguity about how they're to be construed. Writer clarify using instance of indecisiveness as: consider the term lender and two of the feelings that may be found in Webster's New College Dictionary:
a) The rising ground highlighting a river, water, or ocean...

b) An institution for the detention, loan trade, or problem of cash, for the expansion of credit, and for assisting the transmitting of funds.

The job of disambiguation is to ascertain which of the senses of an equivocal word is invoked in a special use of the term. That is performed by taking a look at the circumstance of the word's use. This is the way the difficulty has generally been interpreted in the term sense disambiguation books. A term is thought to possess a limited quantity of distinct senses, regularly distributed with a thesaurus, database, or additional research supply, and the job of the application would be to create a compelled option between these senses for the significance of each use of an equivocal word, depending on the framework of use.

Methodological Preliminaries:

Author describes several important methodological issues come up in the context of word sense disambiguation. They are of general relevance to NLP, but have received special attention in this context. These are: supervised vs. unsupervised learning.

Supervised Disambiguation:

In supervised disambiguation, a disambiguated corpus is readily available for instruction. There's a instruction established of exemplars where each event of the equivocal term t is provided by informative records with a semantic tag. This setting makes supervised disambiguation an instance of statistical classification. The endeavor would be to construct a classifier which properly classifies new instances depending on their circumstance. Pursuing two of the several supervised algorithms utilized to phrase sense disambiguation that display two essential theoretical methods in record language processing: Bayesian group (the formula suggested by Gale et al. (1992) and Details Concept (the algorithm proposed by Brown et al. (1991). Additionally they attest that quite distinct sources of advice could be applied effectively for disambiguation. The first strategy treats the framework of as a carrier of words event without arrangement, but information is integrated by it from several words within the context window. The 2nd strategy appears at just one educational attribute within the circumstance, which might be delicate to text construction. But this feature is carefully selected from a large number of potential ‘informants.’

Dictionary - Based Disambiguation:
We may fall back on a broad portrayal of the feelings. If we don't have any details about the sense categorization of particular examples of a term. This area describes disambiguation procedures that depend on the meaning of feelings in thesauri and dictionaries. Three different kinds of details have been utilized. Lesk (1986) uses the feeling meanings in the book right. Yarowsky (1992) demonstrates how you can employ the semantic categorization of phrases (derived in the classes in Roget's thesaurus) for the semantic categorization and disambiguation of contexts.

Unsupervised Disambiguation:

A benefit of unsupervised disambiguation is that it may be readily modified to create variations between use types that are far more fine-grained than could be present in a book. Again, info access is a software that that is advantageous. The distinction between physical banks in the framework of banks and bank robberies as intangible companies in the context of company mergers may be extremely important even when it's not mirrored in dictionaries.

The authors explain all detail about natural language processing which gives concept about word sense extraction. They depict idea about mapping multiple meaning of word in dictionary. It also gives idea about approaches used in word sense extraction.

5.2.1 The WordNet Model

Because of insufficient understanding cleverness in pc, extra resources are used by word sense extraction to sense word like dictionaries, tagged files etc. WordNet is a broadly employed lexicon for the Language vocabulary. WordNet is likely the very best known lexical database utilized in Normal Language Processing. It organizations Language phrases into sets of synonyms called synsets, supplies brief, basic meanings, and records the numerous semantic relationships between these synsets. In WordNet information is arranged in the conceptual significance level. As a synset which is subsequently associated with additional ideas through associations each notion is known. Word feeling in WordNet is created by team of synset, where a synset of the term w is some w's synonyms. In WordNet each term goes to 1 or more synsets as polysemic term Known.

Each synset also includes a description, called shine, to more explain the feeling that it signifies. Some synsets also have test phrases. Various senses of the term are indicated through synsets. It is called monosemous, otherwise it is
polysemous, if just one synset is accessible for a term. Associations between synsets represent yet another significant element of WordNet. There are many kinds of associations, including hypernyms, hyponyms, holonyms, meronyms, troponyms, coordinate conditions, and thus on. Additionally, associations between phrases may also be potential, the finest instance being antonymy. The relationship is symbolized by synsets themselves. Due to their accessibility, most scientists in Term Sense Extraction function are now counting on the sense distinctions supplied by recognized lexical sources, for example machine-readable dictionaries or WordNet. The dominating design in these studies is the "bank design," which tries to expand the obvious delineation between bank-money and bank-riverside to all sense distinctions. Nevertheless, it's obvious this convenient delineation is in no way appropriate to all as well as most additional phrases.

5.2.2 Problem Definition

Today, unstructured text accessible in the shape of emails, ordinary text file and HTML documents etc. When meanings with regard to circumstance, have to be removed, the disambiguation of contextual multiple connotations of those texts is a complex problem which problem becomes more critical. The research work proposes that Given a sample XML file as an input, word sense on given xml file should be performed with respect to its context. First of all, unstructured input file must be converted into structured format using preprocessing steps of text mining so that correct keyword file is generated. Then by using this keyword file, system will map WordNet relation with Neighborhood Keywords of user inputted word.

5.2.3 Goal and Objectives

Word sense disambiguation is the process of identifying word meanings in a context. The result of this process largely depends on the ability of a system to uncover various semantic relations among words in a text. The objective of the research work is to interpret word sense depending on its context. For this idea of mapping words relation is used. Following are main tasks performed.

i) Capturing words from unstructured text and mapping relation of words.

ii) Deciding the particular meaning of words in a context with different perspective.

iii) To improve the accessibility and usability of text.
5.2.4 Principles

In modern technology and considering feature of information technology, electronic documents are becoming a principle media for human to perform his/her day to day work in industry, educational system, research system etc. In order to fully utilize these online documents effectively, it is crucial to be able to extract the relevant of these documents. Thanks to enormous quantities of unstructured digital texts accessible, automated techniques are required by it to examine and take out understanding from these databases of info. Knowing connotations and semantics of the texts is a complex issue. When meanings with regard to circumstance, need to be pulled, this dilemma becomes more critical. Having Word sense system would be always useful for information retrieval and extraction. Word sense program may decide which sense of the word is employed in a specified framework.

5.2.5 Approaches to Word Sense Extraction

Following are principal strategies utilized in pc to perception phrase.

I. Dictionary-based Algorithms:

It utilizes information assets in the type of machine readable dictionaries to remove multiple feeling of term. Book identifies a phrase in a special area.

Advantages: unnecessary time is saved by it needed to produce training data.

Disadvantages: In this strategy all tagged sense doesn't be provided by it in one book, therefore method is required by it to include new sense in book.

II. Supervised Disambiguation Algorithms:

It utilizes information source in the shape of labeled corpora which identifies meaning of term. It builds classifier which groups new term right based on their context useful. Large sense training is needed by it established to remove sense of term.

Advantages: compared to additional strategies supervised algorithms provide the best efficiency in terms of correctness which may achieve quite nearer to sense removal.

Disadvantages:

Supervised formula needs a substantial number of training information which requires more period to produce it.

III. Unsupervised Disambiguation Algorithm
Unsupervised disambiguation criteria is comparable to clustering where they group examples of the term by significance.

Advantages: Unsupervised algorithms uses training dataset for which marked sense opinion are not needed. Compared to supervised algorithm efficiency is higher.

Disadvantages: Compared to monitor algorithm poor accuracy is given by it.

5.2.6 Benefits of Proposed Research Work

Word Sense Removal is among the jobs in Natural Language Processing. This method addresses two qualities of the British language (and of every other voiced, or "natural", vocabulary for that matter), Polysemy and homonymy. These qualities define two distinct getting exactly the same spelling, and a term that has several feelings, respectively. It has several programs in computational linguistics, including machine translation, info access, text exploration or presentation running. One issue with word sense disambiguation is that it doesn’t have a direct application in human language technology. Indeed, Word Sense Extraction proves its value when integrated into other applications like information retrieval, machine translation and word processing. However, the accuracy obtained by Word Sense Extraction systems until now, is enough to be used for most applications. Therefore, an improvement in the accuracy of word sense extraction would make it much more useful. For example, let’s look at the case of machine translation. In fact, word sense extraction was originally proposed as a subtask of machine translation. In the 1950s, scientists argued that advancements in the former (Word Sense Extraction) will help improve the latter task (MT). Research in both tasks struggles to achieve an acceptable accuracy.

Another field where Word Sense Extraction has a high potential for applicability is the semantic web, where Word Sense Extraction has been used for tasks like ontology learning, building taxonomies and enriching large-scale semantic networks. The semantic web, considered by some the future of the Internet, needs accurate word sense Word Sense Extraction solution in order to evolve. Not only the future can benefit from Word Sense Extraction: Information Retrieval, the currently predominant set of techniques for finding documents on the Internet, would achieve better accuracy, should it use Word Sense Extraction to distinguish the senses of a word that a user searches for, and return only documents containing those senses that are needed. Another issue related to the applications in which Word Sense Extraction can be employed is the method of evaluation.
5.2.7 Research and Development Challenges

Today word sense extraction is extremely important as it found multiple meanings of a same word when context changes. End user required memorizing all the meanings of word in document by means of context which is difficult task for human. Accuracy in correct word sense will retrieve better information from dataset. Currently there are three approaches available to sense word in text documents:

a) Dictionary based approach: It uses dictionaries and lexical knowledge bases to extract exact definition of term.

b) Supervised methods: It uses manually sense-tagged training data to sense word.

c) Unsupervised methods: Currently this method is challenging in the field of word sense extraction from text documents. It groups word on the basis of context similarity. Its performance is very low as compared to other method of word sense extraction from text documents.

In dictionary based approach most of applications currently uses WordNet dictionary to sense word. Synonyms groups signify the component of WordNet: they are a group of words to create a feeling of a specified term. Relationships between synonyms are another important part of WordNet. Briefly, here are some of the challenges on which the research hopes to target focus:

a) What are the important features of a text mining system that extracts the main ideas from original documents?

b) How do we address ambiguous words in the original documents, if any?

c) How should we evaluate the word sense system?

5.2.8 Overview of the Solution

For circumstance established word sense the current study function concentrates on, Pre-Processing on text and removing useful sense from text. In pre-processing stage it requires input signal as Reuter's group of text files. Then it examines the construction of the writing. It supplies mined info to alternative and works cleaning procedure on unstructured text. In removing perception of provided term WordNet dictionary is used by system. WordNet dictionary aids to chart the connection of term in a Reuters files. For preprocessing system work in following step to get result.

a) System first parses xml data files in text files.

b) Using these text data file system extracts important keywords by removing common words.
For extracting exact word sense from text, for e.g. “bank” will map sense in WordNet dictionary as follows.

a) System accept user query in the form of word. For example if user wants to search word sense of “bank” with respect to financial document, then sense should be respect to financial.

b) Here system Extract all important Neighborhood Keywords related to bank like “savings”, “deposit”, “Investment”, “paying” etc. From same document and store it in temporary buffer.

c) Then system extracts all synonyms from WordNet dictionary of bank and compared with Neighborhood Keywords of “bank” until match is not found. Whenever synonym matches with Neighborhood Keywords system will map corresponding sense relation for user input.

5.2.9 Applying Quality Assurance

The quality assurance has been done at three different phases of SDLC:

i) Quality Assurance at Requirements Analysis Phase

ii) Quality Assurance at Design Phase

iii) Quality Assurance at Implementation Phase

And more on following tasks will be performed at each level.

a) Traceability Analysis:

Traceability analysis is an important task to be carried out because it provides the traceability of design and code to the functional requirements. Generally, traceability matrixes can be ready displaying the traceability of signal and style to the practical demands of the program. Functional requirements then as different lines it may consider practical requirements as different posts and design quests and then 1 should have a look at whether for the demand special design component is accessible or maybe not - when the traceability matrix is of design. Merely leave it vacant if yes then place a mark over there otherwise.

b) Checklist Document

The checklist document contains the type of questionnaire regarding different issues focusing on different components at three important phases of SDLC, namely requirements analysis, design and implementation.
5.3 SYSTEM ANALYSIS

Systems evaluation could be the study of models of communicating organizations, including pc systems evaluation. This discipline is closely associated with needs evaluation or operations research. It is also "an explicit formal inquiry carried out to help someone (referred to as the decision maker) identify a better course of action and make a better decision than he might otherwise have made."

5.3.1 System Architecture:

Word sense is definitely an essential issue in removal and information access, too as, text mining, because devices don't have that much cleverness when compared with individual to feeling term in special circumstance. Individual has an capability to discover relationship understanding of term in a circumstance. For instance for a term "Fine" within the circumstance of individual situation it connect more word like "look", "nicely", "feel". Because of these related term human condition won't be definitely described by it rather than send to punishment. As stated above problem, application need to find meaning of a particular word in different context by mapping relation with its neighbor keywords and WordNet dictionary synonym, arranged according to their relevance. Application uses Dictionary-based approach to sense word. Knowledge resources are used by it in the kind of WordNet thesaurus to remove multiple feeling of term. Book identifies a phrase in a special area. As word sense design split into two quests Proven in number context established.

I. Pre-Processing

II. Extracting useful perception from Reuter dataset.

In pre-processing stage it requires input signal as XML document which includes text files. Then it examines the construction of the writing. It supplies mined information to alternative and works cleaning procedure on unstructured text. For removing significant key words it evaluate with quit term checklist and according consequence of assessment, key word document is created by disregarding frequent phrases. In second module, user will input word for sensing its meaning with respect to keyword generated in previous step. WordNet dictionary is used by system, to take out feeling of provided word sense. WordNet dictionary aids to chart the connection of term in a Reuter's files. The main purpose of second module is to extract word sense according to its context. For this, idea of dictionary synonyms and neighboring words mapping relation is used. After matching dictionary synonyms and neighboring words relation,
it extracts relevant sense from dictionary. As shown in figure 5.1 user will input word to be sensed. System performs pre-processing on documents and generates keyword files. System will group Neighboring Keyword of inputted word from documents. Grouped Neighboring Keyword will compare with synonym from dictionary and if matches it will display sense of inputted word. For example consider following two documents for word “bank”.

“A trust company is a corporation, especially a commercial bank, organized to perform the fiduciary of trusts and agencies. It is normally owned by one of three types of structures: an independent partnership, a bank, or a law firm, each of which specializes in being a trustee of various kinds of trusts and in managing estates. Trust companies are not required to exercise all of the powers that they are granted.”

“A Bank account is a financial account recording the financial transactions between the customer and the bank and the resulting financial position of the customer with the bank. Bank accounts may have a positive, or Tax paying balance, where the bank owes money to the customer.”

In above example, Group the Neighboring Keywords for the word bank.

<table>
<thead>
<tr>
<th>Bank</th>
<th>trust</th>
</tr>
</thead>
<tbody>
<tr>
<td>account</td>
<td>company</td>
</tr>
<tr>
<td>financial</td>
<td>corporation</td>
</tr>
<tr>
<td>account</td>
<td>commercial</td>
</tr>
<tr>
<td>recording</td>
<td>bank</td>
</tr>
<tr>
<td>financial</td>
<td></td>
</tr>
<tr>
<td>transactions</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Group of Neighboring Keywords for the word ‘bank’

Here mapping relation synonym with trust, financial, account etc. then it’s corresponding sense is:

I. Trust, swear, rely, bank -- Have confidence or faith in.

II. Deposit, bank, financial, account -- Put into a bank account.
Figure 5.3: Architecture of Word Sense System based on its context
The research work is accomplished in three different modules as under.

i) Pre-Requisite Project Management Tasks.

ii) Primary Project Management Tasks.

iii) Supportive Project Management Tasks.

5.3.2 Pre-Requisite System Management Tasks

The pre-requisite system management tasks are those tasks which are required prior to start with the development of the system. One of task is requirements gathering which obtains the requirements of a system from users, research papers.

The different tasks in requirements gathering are:

- Gathering and reading documents regarding previous work-done.
- Review existing systems or frameworks
- Interview different people who are supposed to use the system or software
- Sort out the requirements

For this work several research papers had been studied regarding word sense system and text mining. Idea for sensing word is developed by relating its associated word to a dictionary sense.

5.3.3 Primary System Management Tasks

Primary system management tasks are those tasks which can be called as the core of the system. These tasks include:

i) Concept Development

ii) Analysis and Design

iii) Implementation (Coding)

iv) Testing (Program)

i) Concept Development

The Idea Development stage is the start of the study function. Within this period, the notion for that program is elaborated and investigated. The aim of this period would be to analyze the feasibility of the program [47]. Within this period, the present or future system leader creates a suggestion, which comprises a outline of the above-mentioned issues. Types of the kind of study proposal contain company ideas and grant programs. Why this system? : concerns to be clarified within the Theory Advancement are these system?

Is it achievable?
Who are potential companions in this system?

What if the results be?

What are the borders of the program (what is outside the range of the program)?

ii) Analysis and Design:
The evaluation stage identifies the requirements of the machine, independent of how these needs will undoubtedly be achieved. This stage identifies the difficulty that the client is attempting to resolve. The outcome by the end of the period is really a necessity record. Preferably, this record says in an exact and clear style what's to be assembled. This investigation signifies the "what" period. The necessity document attempts to fully capture the conditions from the customer's view by determining objectives and relationships in a degree eliminated from the execution details [48].

Within the look stage the structures is made. This phase begins with the requirement report shipped from the necessity stage and roadmaps the needs into structures. The buildings identifies the actions, their interfaces and parts. The deliverable style file is the structures. A plan is described by the design document to apply the demands. This stage signifies the "how" stage. Details on computer-programming languages and conditions, machines, programs, program architecture, distributed architecture adding, memory dimensions, platform, calculations, information structures, international sort definitions, interfaces, and a number of other architectural details are proven. The layout can contain the use of current parts.

iii) Execution:
The execution strategy creates the required assets and routine. It identifies execution details including development languages, systems, programming environments, debuggers and additional. Implementation is the actual execution of the plan, design, model or specification. In computer science field it can be termed as the actual program for the algorithm developed that can be executed on any computer system in the world despite of its underlying architecture. Many implementations may exist for the same algorithm as there are so many programming languages available such as JAVA, .NET, VB, ASP, PHP, PERL etc. and many more. In IT industry there would be basically two kinds of concern: User concerns and Designer concerns. Examples of both of these concerns are as follows:

User Concerns:

• Will the system deliver the information required?
• How quickly the data can be accessed?
• How easily the data can be retrieved?
• How much clerical support will be required to enter data into the system?

Designer Concerns:
• How many lines of program code will it take to perform this function?
• How can there be cut down on CPU time when we run the system?
• What are the most efficient ways of storing this data?

iv) Testing:

The check strategy identifies the screening needed to create quality for that program. Then it's announced to be entire, when the machine moves all assessments within the check strategy. It's regarded as of top quality when all tests does be passed by the system then. The more total the protection of the program, the greater is the assurance in the program: consequently the system's quality increases. As the first achievement in the screening period within the layout, that is the place obtained here or it, might be looked at the check strategy could be considered. Among the aims of the layout period, is to create a strategy to finish the program, consequently it's quite organic to incorporate the check strategy. Additionally the trade-offs between alternative architectures may be affected by variations in their check ideas. A single check may workout just a part of the machine. The protection of the check is the percent of the program practiced by the check. The protection of a collection of assessments is the marriage of the coverage of every individual check in the collection.

Generally, it's adequate that the check collection do not need to protect the whole program and comprises all the atypical and typical situations. Reasonable quality is given by this for that expense of assets. All the normal and atypical situations have to be protected, but in doing this, maybe not all posts of delivery inside the machine might be protected. The machine might include internal divisions, mistakes, or stops that will result in untested strings of delivery. Program screening may be said while the procedure of validating and confirming the program created earlier in the day in execution stage and supply subsequent information about:

• Whether it meets the requirements that guided its design and development; and
• Whether it works as expected.
5.3.4 Supportive System Management Tasks

Supportive system administration job is the Quality Assurance actions at Needs Design, Evaluation and Execution stages of the SDLC. Software quality guarantee is an umbrella exercise that is employed at each part of the software procedure. SQA encompasses procedures for the successful program of resources and approaches, formal technical evaluations, testing methods and methods, procedures for change manage, procedures for ensuring conformity to requirements, and confirming and measurement systems. Computer software evaluations are one of the more significant SQA activities. Evaluations function as filters during all computer software architectural activities, removing mistakes while they're comparatively cheap to proper and find.

Quality assurance is an important task to carry out to ensure the quality of the work done. In the research work quality assurance activities performed at Requirements Analysis, Design and Implementation phases of the SDLC. For this checklist document is prepared which consists of the questionnaire regarding different important issues at all the three phases.

5.4 REQUIREMENT ANALYSIS

Requirements analysis includes those jobs that go in to determining the demands or conditions to match for a new or adjusted system, taking accounts of the potentially conflicting needs of the numerous stakeholders. Requirements are the desired characteristics of the software being developed. The first activity in most projects is the identification and documentation of the requirements. Requirements cover both requirements engineering i.e. identification, analysis and capture and requirements management which contains managing change, creating and maintaining agreement with customers, trace ability and metrics.

Many challenges are presented by the development of large, complex systems to systems technicians. Foremost among these could be the skill to make sure that the ultimate program fulfills the requirements of customers and offer simple care and augmentation of these methods in their used life. These methods frequently change and develop during their life period. This makes it hard to monitor the program from the initial and changing consumer needs. Conditions create an awareness of user's require as well as supply the last yardstick against which implementation achievement is calculated. Various studies have shown that roughly half of the application errors
can be traced to requirement errors and deficiencies. Thorough documentation and properly managing requirements are the keys to developing quality applications. The following requirements and constraints were considered during the requirement analysis phase. Conditions could be classified as Practical, Behavioral, New and Nonfunctional.

5.4.1 Architectural Requirements:
- The system components should work in accordance with each other.
- All the components should be placed in one architectural environment.

5.4.2 Behavioral Requirements:
- All the components should coordinate their work properly.
- Functions should respond to the users’ queries properly.
- There should not be confusing operations on users’ end.
- All tasks to be performed with ease without giving much more stress on users’ side.

5.4.3 Functional Specification
Practical Standards is the initial period of Software Development Life Cycle. High quality assurance much more depends on this phase because it gives you defects, traceability, verification and validation in the early phase of system development. Even if implementation is not started still you can review each module of the system by preparing quality assurance checklist document. Both development and quality team involved at this phase to perform quality assurance. Following functional standards prepared for circumstance established word sense removal in text.

a) Pre-requisites:
Before using context based word sense extraction in text following resources are required.
- Installed WordNet Dictionary of minimum version 2.1 for sensing a word
- Reuter 21578 Dataset and its extracted important keywords
- Installed Java Environment for running system successfully

b) Input Requirements:
- Reuter 21578 XML Dataset
- Reuter 21578 Parsed XML Files
- Reuter 21578 Extracted Keywords
- Word entered by user sense its meaning

c) Output Requirements:
- Reuter 21578 Parsed XML Files
- Reuter 21578 Extracted Keywords
- Word sense according to context

Context based word sense system should
- Have to determine number of sense in a dataset for a given word.
- Able to find out relationship between words.

d) Resource requirements:
For this system input provided must be in text file. As the size of the text document changes, performance of the system also changes. So for larger files better hardware with better performance is required.

Hardware Requirements:
- Processor: Pentium IV or higher
- RAM: Minimum 256 MB
- Hard Disk: 40 GB
- Input device: Standard Keyboard and Mouse
- Output device: VGA and High Resolution Monitor

Software Requirements: Software components required for the system development are:
- Operating System: Windows XP or above
- JDK: Java Development Kit
- Language: JAVA

Tools required: WordNet Dictionary

5.4.4 Non-Functional Specification
a) Developmental Constraints:
The system will depend on the Reuter 21578 Dataset and its extracted important keywords. It uses interface WordNet dictionary to sense word, but it is possible to interface our own tagged corpus.
b) Dataset Constraint:

Within this function particular Reuter 21578 Dataset utilized for executing word feeling on text information. The dataset is limited into xml format. Another data used in this work is WordNet dictionary which gives word sense.

c) Time Constraint:

The system is confined to be completed within specified period. A schedule will work out which made it even more challenging.

d) Performance

- Results from Context based word sense system should give accuracy.
- Performance of the Context based word sense system should satisfy with input query
- Response time should be faster.

5.4.5 Feasibility Study

The proposed system may be not feasible with some components which are used while developing system. Therefore it is necessary to evaluate feasibility of the system at the earliest stage. The feasibility is based on following properties:

a) Platform and Technology:

This type of feasibility contains study of functions, performance, and constraints that may affect the ability to achieve an acceptable system. The system is technically feasible for implementation. No extra hardware is required while developing this system. The software can be installed on a single machine with Windows-XP or a higher version of Operating System.

b) Economical Feasible:

Economical feasibility is the evaluation of the development cost weighed against the ultimate income or benefits derived from the developed system. This project is financially feasible because the software does not require any extra hardware or any additional supporting technology which in turn adds no extra cost to the software. Thus the cost is only for the development and hence system is financially feasible.
c) Resources

Context based word sense system requires only a single or multiple machines. Thus no additional resources are required to implement the system. Thus the software is also resource feasible.

5.4.6 Risk Management

Risk into software system can affect desired goal or can give big loss to product. In order to avoid software risks we have to categorize risk and find a solution to avoid such risk. For the research purpose the risk is classified in two categories.

a) System management risks:

Potential budgetary, schedule, personnel, resource, customer, and requirements problems are included by these risks, and chances are that the program schedule will slide and that prices will grow, when they become actual. An essential danger is program management. Because, software is intangible, it's quite hard to manage and track a software program. The progress of the work is assessed on regular basis to avoid the risk.

b) Platform and Technology risks:

Technologies and system risks project problem, possible layout, execution, interfacing, screening and care difficulties. The standard method is followed so that unambiguous specifications are performed. New things are tried every time that required for the system. Scope and Purpose of Document highlights the various risks that may be encountered while following the software development lifecycle for software development. This document highlights the risks that may affect the smooth working of the system development.

c) Overview of major Risks

The identification of risk is central to the success and failure of a system. A concentrated effort to minimize and even eliminate certain risks related to the system is made. The following list of risk item checklists analyze each risk and related counter effects and policies adopted by the system development.

- Development Environment
- Technical/resource Failures
- Quality checklist for each step
5.4.7 Risk Identification

Threat Id is a methodical effort to pin down risks for the Study function strategy. By determining the foreseeable and known dangers, a first step is taken by the project manager toward preventing them when feasible and managing them when needed [38]. One approach for determining dangers is to produce a risk item record which may be utilized for risk recognition. Risk item checklist is prepared on following subcategories: product size, Technical approach and technology to be built.

<table>
<thead>
<tr>
<th>Sr.No.</th>
<th>Risk Category/ Risk item</th>
<th>Risk Description</th>
<th>Risk Checklist Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Product Size Risks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>Input size</td>
<td>Maximum input file size used</td>
<td>1.23 MB</td>
</tr>
<tr>
<td>b</td>
<td>Program Size</td>
<td>Estimated size of the product in Line of Code.</td>
<td>900 LOC</td>
</tr>
<tr>
<td>c</td>
<td>Output Size</td>
<td>Maximum Output size file</td>
<td>750 KB</td>
</tr>
<tr>
<td>II</td>
<td>Technical risk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>Software requirement analysis</td>
<td>What is specific output by software requirement analysis?</td>
<td>Functional and non-Functional specification</td>
</tr>
<tr>
<td>b</td>
<td>Software planning</td>
<td>What specific method used for planning the system?</td>
<td>System Schedule chart</td>
</tr>
<tr>
<td>c</td>
<td>Software Quality</td>
<td>How Software quality is achieved in Development process?</td>
<td>Output Assessment of each step</td>
</tr>
<tr>
<td>III</td>
<td>Technology Risk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Technology</td>
<td>Are you new to development technology?</td>
<td>No</td>
</tr>
</tbody>
</table>
2 Interface Is there Interface necessity between hardware and software? No.

3 Compatibility Is system compatible with platform? Yes

<table>
<thead>
<tr>
<th>Table 5.2: Risk Identification Checklist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface</td>
</tr>
<tr>
<td>Compatibility</td>
</tr>
</tbody>
</table>

5.4.8 Risk Projection

Risk projection, also called risk appraisal, tries to price each risk in two ways—the chance or probability that the risk is actual and the outcomes of the difficulties connected with all the risk, should it happen. As shown in table 5.3 risk projection is prepared for the system. First column describes risk summary that we analyzed in system. Each hazard is classified within the 2nd line. The chance of incidence of each hazard is joined in the next line of the table. Next, the effect of every hazard is evaluated. According to requirements risk impact described in four categories the following.

i) Devastating
ii) Crucial
iii) Minor
iv) Minimal

The categories for every of the four risk parts operation, help, price and routine are averaged to decide an general impact worth.
<table>
<thead>
<tr>
<th>Risk Description</th>
<th>Risk Category</th>
<th>Risk Probability</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirement not clear</td>
<td>System management</td>
<td>10%</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>risks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>As per schedule project completed in time</td>
<td>System management</td>
<td>10%</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>risks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sufficient budget not available</td>
<td>System management</td>
<td>5%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>risks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected Resource not available</td>
<td>System management</td>
<td>15%</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>risks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training on used tool not sufficient</td>
<td>Platform and Technology risks</td>
<td>15%</td>
<td>4</td>
</tr>
<tr>
<td>Technology not comfortable with system</td>
<td>Platform and Technology risks</td>
<td>5%</td>
<td>1</td>
</tr>
<tr>
<td>System is not scalable</td>
<td>Platform and Technology risks</td>
<td>5%</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.3: Risk Projection

5.4.9 Risk Monitoring and Mitigation

These risks are given in the above table can be monitored in the following ways to restrict their involvement and thus the system failure can be reduced.

i) As per schedule system completed in time

This risk can be eliminated by the formulation of a system plan, which can be checked by taking desk review time to time.

ii) Requirement not clear

This risk can be eliminated by preparing questionnaire depending on market survey.

iii) Technology not comfortable with system

This risk can be prevented by careful analysis of the system requirements and feasibility before coding starts.

iv) Training on used tool not sufficient

This is a risk that is due to inexperience of the developers who developing the software. To remove this risk, guidance from expert programmers should be taken.
v) Expected Resource not available

This risk can be prevented by availing clear requirement of the resources.

vi) Insufficient response

The response time of the system is poor because of low memory and low configurations. This risk is to be mitigated by building high quality software by optimizing resource usage in order to the performance.

vii) Lack of Experienced and trained developer

This is not a very critical risk, since the software knowledge can be gained and hardwork can overcome this problem.

5.4.10 Planning and Scheduling

Scheduling of the study function doesn't vary considerably from arranging of any multitasking attempt. Therefore generalized program booking tools and methods could be employed to analyze use small changes. The program program offers a road chart for study function improvement. The program program identifies because the study work carry on the goals and duties that have to be monitored and managed, if it's been correctly created. Problem statement is solved by using following four phases.

5.4.10.1 An Input Phase

An input phase in which a problem is detected and an attempt is made to understand the problem. Human language is ambiguous, so that same word can be interpreted in multiple ways depending on the context in which it occurs. Problem states that extract word sense according to its context. Proposed system performs easier task for end user to determine which sense of a word should be adopted for each instance of a word. It will decide best meaning of words in a context with different perspective. Meaning is decided by mapping keyword relation with dictionary synonyms.

5.4.10.2 A Processing Phase:

A processing phase in which problem statement is divided into modules and planning is done for evaluating solution for problem statement. Provided issue declaration for term sense program depending on its framework is split in to two quests.

1. Pre-Processing

2. Extracting useful perception from Reuter dataset
In pre-processing stage it requires input signal as XML record. Then it examines the construction of the writing. It supplies mined info to more measure and works cleaning procedure on unstructured text. It provides all extracted keyword to next step. In extracting word sense module application uses WordNet dictionary to sense word. WordNet dictionary aids to chart the alternatives relationship with nearby words in document.

5.4.10.3 An Output Phase:

An output phase in which includes solution to derived problem in earlier phase. It evaluates solution by converting design into application by computer language. This phase gives output to find solution for problem statement.

5.4.10.4 A Review Phase

If demanded, an evaluation stage where the answer is assessed and modification is created. A review is applied to main phases of the research work development in the form of quality assurance. Quality checklist performed in order to get expected review. Table 5.4 shows activity chart which contain task, duration, begin date, and End date following activity. Figure 5.4 shows sample Gantt chart.

<table>
<thead>
<tr>
<th>ID</th>
<th>Task Name</th>
<th>Begin date</th>
<th>End date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Information Gathering</td>
<td>5/2/2012</td>
<td>7/3/2012</td>
</tr>
<tr>
<td>3</td>
<td>Concept Development</td>
<td>27/4/2012</td>
<td>5/5/2012</td>
</tr>
<tr>
<td>4</td>
<td>Planning and Scheduling</td>
<td>6/5/2012</td>
<td>9/5/2012</td>
</tr>
<tr>
<td>5</td>
<td>Requirements analysis</td>
<td>11/5/2012</td>
<td>23/5/2012</td>
</tr>
<tr>
<td>6</td>
<td>Risk identification and monitoring</td>
<td>25/5/2012</td>
<td>5/6/2012</td>
</tr>
<tr>
<td>7</td>
<td>Design and modeling</td>
<td>6/6/2012</td>
<td>24/6/2012</td>
</tr>
<tr>
<td>8</td>
<td>Design review and refinement</td>
<td>25/6/2012</td>
<td>2/7/2012</td>
</tr>
<tr>
<td>9</td>
<td>GUI design</td>
<td>3/7/2012</td>
<td>15/7/2012</td>
</tr>
<tr>
<td>10</td>
<td>Implementation</td>
<td>17/7/2012</td>
<td>29/7/2012</td>
</tr>
<tr>
<td>11</td>
<td>Review and suggestions for Implementation</td>
<td>30/7/2012</td>
<td>31/8/2012</td>
</tr>
<tr>
<td>12</td>
<td>Testing and Quality Assurance</td>
<td>1/9/2012</td>
<td>28/9/2012</td>
</tr>
<tr>
<td>13</td>
<td>Review and suggestions for Testing and QA</td>
<td>29/9/2012</td>
<td>12/10/2012</td>
</tr>
<tr>
<td>14</td>
<td>Refined QA activities</td>
<td>13/10/2012</td>
<td>12/11/2012</td>
</tr>
</tbody>
</table>

Table 5.4: Activity Chart for Context based word sense system
5.4.10.5 System Cost Estimation

COCOMO is a comparatively simple design according to advices associated with numerous price motorists along with how big the machine productivity is affected by that. His co-workers and Boehm have identified and updated COCOMO, the outcome is called COCOMO that is the reason current modifications in computer software engineering technologies.

Figure 5.4: Sample Gantt chart
The Price Estimate COCOMO model is made for three groups of software program:

I. Organic Manner
Fairly small, straightforward software program by which little group with great program expertise function to some group of less than firm specifications.

II. Semidetached Mode
An advanced dimension and sophistication of computer software program by which group with combined experience function should fulfill a combination of rigid and less than rigid specifications.

III. Embedded Setting
A computer software program that has to be created within an operational limitations.

Table 5.5 shows constant values of constant for cost estimation mode.

<table>
<thead>
<tr>
<th>Mode</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organic</td>
<td>3.2</td>
<td>1.05</td>
</tr>
<tr>
<td>Semi Detached</td>
<td>3.0</td>
<td>1.12</td>
</tr>
<tr>
<td>Organic</td>
<td>2.8</td>
<td>1.20</td>
</tr>
</tbody>
</table>

Table 5.5: Value of constants for cost estimation mode

\[ E_i = A \times (KDLCO)^B \]

Where,
- \( E_i \) = initial effort
- \( A, B \) = Constant depending on the project being organic semi detached or embedded
- KDLOC = deliverable thousand line of code

\[ E = EAF \times E_i \]

Since, Organic Model is used,

\[ E_i = A \times (KDLCO)^B \]

= 3.2 (KDLOC) ^ 1.05
= 3.2 (2) ^ 1.05
= 7.02
E = 0.9537 * 7.02
= 6.69PM

5.4.11 DFD
DFDs do not provide comprehensive explanations of quests but graphically explain a system's data and how the data connect to the program. Information Circulation Diagrams serves two purposes: a)To supply an indicator of how data are changed as they go through the program. b)To illustrate the capabilities that changes the data flow. The DFD offers added information that's utilized throughout the evaluation of the information domain name and functions being a foundation for your modeling of perform. A explanation for every function offered in the DFD is comprised in a procedure standards. The information flow plans can be utilized to signify something or applications at any degree of abstraction. In truth, DFDs might be portioned in to amounts that signify growing info circulation and practical depth.

Notations for DFD are:
(a) External Entity   (b) Basic Process (c) Individual Process (d) Interface (e) Data store (f) Flow

Fig. 5.5 shows the level 0 DFD for the ‘Word Sense text-mining system based on its context’. In this diagram, user acts as external entity, Reuter dataset act as data-store which is used for retrieving neighborhood keyword of word and ‘Word Sense text-mining system based on its context’ acts as basic process for extracting possible sense for word in a dataset.
Fig. 5.5 shows the level 1 DFD for the ‘Word Sense text-mining system based on its context’. The basic process is broken down into sub-processes such as ‘Process 1: Parsing, ordering and finding key-phrase’, ‘Process 2: Accepting user query’, ‘Process 3: Select meaning for word’, ‘Process 4: Display associated word senses’, Here KeyPharase text miner, system interface act as interfaces which operate various processes.
5.5 SYSTEM MODELING AND DESIGN

A design captures aspects significant for many programs while omitting (or abstracting) the remainder. A design within the circumstance of software development may be visual, textual, numerical, or application code-based. Versions are quite helpful in recording evaluation and style outcomes. Versions additionally ease the style and evaluation processes themselves. Visual versions are quite well-liked simply because they are simple to build and comprehend. Single Modeling Language (UML) is a general objective visual modeling language that is utilized to define, visualize, assemble and record the items of the computer software system. UML is chiefly a graphic modeling tool. Nevertheless, it frequently needs the graphical models to be accompanied by text explanations. UML may be used to assemble eight different kinds of blueprints to get five different perspectives of the program. The various UML diagrams offer different views of the computer software program to be created and help a complete knowledge of the machine. Such versions could be processed to obtain the real execution of the device.

5.5.1 Use Case Diagram

The goal of a use situation would be to determine an item of defined behaviour without displaying the inner arrangement of the program. The use cases don't mention any particular formula to be utilized or the internal information representation, internal arrangement of the applications, and so on. A use situation commonly signifies a series of relationships between the program and the person. Use case plans replies what program does in the person perspective. Use case solution 'What may the system do?' There are three essential components in a use case scenario, celebrity and use case.

a) Scenario:

A scenario is a series of occasions which occur whenever a person interacts with the system. The important scenario according to the system is as follows:

i) User Inputs XML file for Preprocessing.
ii) System will parse XML file into text file.
iii) Then system will generate Keyword file from parsed text file.
iv) User will click action on display keywords.
v) User inputs keyword file to sense word.
vi) User inputs word to be sensed in keyword file.
vii) System will collect neighborhood keyword of inputted user word.

viii) System will get synonyms of inputted word from WordNet dictionary.

ix) After mapping relation system will display sense of word.

a) Actor

Actor may be Word Sense text-mining system based on its context or end user.

b) Use Case

Use case is job or the objective done from the consumer. An easy situation with 'Actor' and 'Use Case' referred to as uses. Actor's are symbolized by easy stick guy and use situation by oblong form as proven in fig. 5.7.

Use Case Descriptions

Use Case 1: Input XML File
Purpose: Input for preprocessing.
Primary Actor: End User who uses this application.
Preconditions: The system must be started and end user selects XML file.
Flow of Events

• User browse XML file.
• System takes input as XML file.
• Input XML file is given to parser.

Post condition: Parsed file stores data in text file.

Use Case 2: Parse XML file into text file
Purpose: Remove XML tag and extract text body.
Primary Actor: System which extract text.
Preconditions: The system must be given input as XML file.
Flow of Events:

• System gets XML file.
• System Parse XML file by removing tags.
• System stores text data in parsedout text file.

Post condition: Extract keyword from text file.
Use Case 3: Generate Keyword file from text file

Purpose: Remove common words from text file.

Primary Actor: System which extract keyword.

Preconditions: The system must be given input as parsed text file.

Flow of Events:

- System gets parsed text file.
- System compares each token of parsed text file with common words.
- System stores unmatched token in keyword file.
Post condition: Generated Token display on screen.

Use Case 4: Display Keyword on screen
Purpose: To view extracted keyword result.
Primary Actor: End user who uses this application.
Preconditions: The system must create parsed keyword file.
Flow of Events
- System stores extracted keyword in keyword file.
- System gets keyword file.
- System display keyword on screen.
Post condition: Extracted keyword display on screen.

Use Case 5: Input Keyword file
Purpose: To search keyword in file.
Primary Actor: End user who uses this application.
Preconditions: The system must create parsed keyword file.
Flow of Events
- User Clicks on word sense.
- System gets input as keyword file.
Post condition: User inputs word to be sense.

Use Case 6: Input word to be sense
Purpose: Interpret word sense.
Primary Actor: End user who uses this application.
Preconditions: The system must inputted parsed keyword file.
Flow of Events
- User inputs word to be sensed.
- User execute query.
- System display sense for given input.
Post condition: System display sense on screen.

Use Case 7: Get neighborhood keyword
Purpose: Find associated word.
Primary Actor: System which computes neighborhood keyword.
Preconditions: The system must have input, parsed keyword file and word to be sensed.
Flow of Events
- System gets keyword file and word to be sensed.
- System search keyword in keyword file.
- System computes neighborhood keyword for input word.

Post condition: neighborhood keyword stored in buffer.

Use Case 8: Get synonyms from dictionary
Purpose: Map Synonyms with neighborhood keyword.
Primary Actor: System which computes synonyms from dictionary.
Preconditions: The system must have input, parsed keyword file and word to be sensed.
Flow of Events
- System gets keyword name.
- System extract dictionary synonym for given keyword.
Post condition: System stores synonym in buffer.

Use Case 9: Display word sense
Purpose: Display sense for input word.
Primary Actor: End user who uses this application.
Preconditions: The system must have input dictionary synonym, neighborhood keyword and word to be sensed.
Flow of Events
- System gets input as keyword.
- System map dictionary Synonyms with neighborhood keyword.
- If match found it display sense on screen.
Post condition: System display sense on screen.

5.5.2 Sequence Diagram
Interaction is shown by a sequence diagram among like a two dimensional graph things. The graph is examine all the way through. The vertical dashed point is known as the item's lifeline. The lifeline signifies the lifestyle of the item at any special level of time. The rectangle drawn on the life is called the service image and signifies that the item is effective so long as the rectangle exists. As an arrow between the lifelines of two items each information is suggested. The emails are shown in date sequence from the best towards the underside. That's, studying the plans from the best to the
base might display the series by which the communications happen. Each message is
tagged using the message title. Some handle info may also be contained. Two kinds of
manage info are especially precious.

- A condition indicates that a message is sent, only if the condition is true.
- An iteration marker shows the message is sent many times to multiple receiver
  objects.

Figure 5.8 shows 'Sequence diagram' of the system. In this sequence diagram there are
three objects 'System Interface', 'Keyword Generator' and 'Word Sense System'. The
information circulation is shown vertically in waterfall fashion i.e. it begins from the
best and moves towards the underside. Dashed outlines signify the length that the item
may be reside. Flat rectangles on the dashed outlines signify service of the item.
Communications delivered from an item are displayed by dark arrow and dark arrow
head. Return information are symbolized by filled arrow. Therefore the number 5.8
exhibits the subsequent series of conversation between your three items:-

User will send browsed XML file to system interface object.

i) System interface object pass XML file to Keyword Generator object.
ii) Keyword Generator object parse XML file into text file.
iii) Then Keyword generator generates keyword file from parsed text file.
iv) Keyword generator object sends message to system interface object to display
    keywords.
v) User will send browsed keyword generated file and word to sense in keyword
    file to word sense system object.
vi) Word sense system map WordNet relation with user inputted word.
vii) Word sense system object sends message to system interface object to display
    sense of inputted word.
5.5.3 Collaboration Diagram

A collaboration diagram exhibits both behavioral and structural aspects explicitly. This really is unlike the behavioural aspects are shown only by a sequence diagram which. The architectural part of a cooperation diagram includes items as well as the links present between them. A effort diagram exhibits the arrangement of emails that execute an procedure or perhaps a trade. Collaboration images reveal things, their hyperlinks, and their emails. Figure 5.9a and 5.9b shows collaboration diagrams for preprocessing and word sense system with sequence respectively.
For Preprocessing

Figure 5.9a Collaboration Diagram for Word Sense System based on its context

For Word sense in text

Figure 5.9b Collaboration Diagram for Word Sense System based on its context
5.5.4 Activity diagram

Activity diagram is still another essential plans in UML to explain powerful areas of the machine. Activity plans is essentially a flow graph to signify the flow type one activity to a different exercise. As a procedure of the machine the action can be described. Therefore the manage stream is attracted in one procedure to a different. This circulation may be consecutive, branched or concurrent. The fundamental goal of exercise plans is to fully capture the dynamic behaviour of the program. As demonstrated in figure 5.10 action diagram splits in two choices by providing consumer options as preprocessing and term sense extraction. For the preprocessing it reads XML file and parse it into text file. From parsed text file it will generate keyword file by removing common words from text file. Finally it displays keyword to the user. In word sense activity it loads first Keyword file and takes user input as word to be sensed. Then it will group neighboring keyword of inputted word and map relation with dictionary synonym of inputted word. Depending on mapping result it will display sense to the user.

Preprocessing Activity:
- User Input File.
- Parse XML file into text file.
- Generate Keyword file from parsed XML file.
- Display Keywords.

Word sense extraction Activity:
- User Input keyword file.
- User input word to sense.
- Search neighborhood keyword of input word.
- Map WordNet relation of neighborhood keyword.
- Display word sense.
Figure 5.10 Activity Diagram for Word Sense System based on its context
5.5.5 Component Diagram

Chief objective of element plans is to exhibit the architectural connections involving the parts of something. Element blueprints will vary when it comes to behaviour and character. Element blueprints are employed to design physical facets of something. Bodily factors are the weather like executables, libraries, documents, records and so on which exists in a node. The following use of element blueprints can be.

- Model the components of a system.
- Model executables of an application.
- Model system's source code.

Figure 5.11 shows different component like Key Phrase text miner, context based word sense miner and user interface etc.

![Component Diagram](image)

Figure 5.11: Component Diagram of Word Sense System based on its context

5.5.6 Deployment Diagram

Deployment plans can be used to explain the fixed deployment see of the program. Figure 5.12 shows Deployment diagram consist of nodes and their relationships.
5.5.7 Algorithm Design

An algorithm is composed of a finite set of steps, each of which may require one or more operations. In computer engineering the word ‘Algorithm’ has a special significance and it has come to refer to a method that can be used by a computer for the solution of a problem. This is what makes algorithm different from terms such as process, technique, or method. The term algorithm is universally used in computer engineering to describe problem-solving methods suitable for implementation as computer programs. When a very large computer program is to be developed, a great deal of effort must go into understanding and defining the problem to be solved, managing its complexity, and decomposing it into smaller subtasks which can be easily implemented.

5.5.7.1 Algorithm for Preprocessing

Preprocessing(String xmlfile):
{
    String xmlfile,parsedout,keyword,strtoken,stoplist;
    // Xmlfile,parsedout,keyword, stoplist all are file names used in preprocessing.
    Message,Input xml file.
    Read(xmlfile);
    StreamBuffer=Parse(xmlfile);
Write StreamBuffer to Parsedout text file.
Write parsedout= StreamBuffer;
Read(parsedout);
ReadStoplist(stoplist)
While(!EOF(parsedout)
{
    Open text file keyword to write tokens.
    Open(keyword);
    While(hasmoretokens)
    {
        ReadTokens(Strtoken);
        ReadStoplist(stoplisttoken);
        If(Strtoken==numeric[0-9])
        {
            Continue;
        } //End of if
        ElseIf(Strtoken==Specialmark["|\+\&\(|\)|\.|\:|'|"]", ",")
        {
            Continue;
        } //End of elseif
        If(Strtoken== stoplisttoken)
        {
            Continue;
        } //End of elseif
        Else
        {
            Write extracted keyword in text file.
            Append Keyword=Strtoken;
        }//End of else
    } //End of inner while
} //End of outer while
} //End of Preprocessing Algorithm
As shown in preprocessing algorithm it takes input as XML file in string variable xmlfile. It uses parsedout file name for writing parsed xml file and using this parsed out file algorithm creates keyword file used in wordsense. As shown in algorithm it removes common words using stoplist file tokens while creating keyword file. It also checks for other tokens.

5.5.7.2 Algorithm for Word Sense

WordSense(keyword):
{
    int wordpos=0,count=0,i;
    String strtoken,senseword,synonym,frontneighbor,backneighbor;
    Input word to be sensed;
    Input(wordsense);
    Synonym=getsynonym(wordsense);
    Read(Keyword) file;
    While(!EOF)
    {
        wordpos++;
        While(hasmoretokens)
        {
            Read(strtoken);
            If(wordsense==strtoken)
            {
                Count=wordpos;
                Repeat for i=count to wordpos + 5
                {
                    If Strcmp(synonym,strtoken(i))
                    {
                        Display word sense from dictionary;
                    }
                    i++;
                }//End for loop
                Repeat for i=count to wordpos - 5
                {
            }
If Strcmp(synonym,strtoken(i))
{
    Display word sense from dictionary;
}
    i--;
}//End for loop

}//End of if
Else
{
    Display word not found;
}
}//End of inner while loop
}//End of outer while loop
}//End of WordSense Algorithm

As shown in wordsense algorithm it will take first input as keyword file and word to be sensed in that file. It extracts all synonyms of word sense from dictionary. Next it searches wordsense in keyword file. If wordsense matches with strtoken, then it will extract front and back keyword from file and check it with synonyms. It will display word sense if it matches synonyms and neighboring keywords.

5.6 IMPLEMENTATION
In computer technology, an execution is a recognition of the technical standards or formula like a course, software part, or additional computer program through installation and encoding. Several implementations might exist for confirmed standards or regular. For instance, web browsers include implementations of Worldwide Web Range. Computer software improvement resources and suggested specs include implementations of programming languages.

5.6.1 Introduction
The style framework established word perception program is needed to be changed into computer knowing language. Design of system is abstract view of system and implementation is real view of system. Implementation phase transfers design of system into computer programming. Execution is the period of the study
function throughout which concept is switched into exercise. In this period developer changes the plan specs into pc guidelines, which called applications. The applications organize the info actions and manage the whole procedure in something.

The aim of the execution period would be to execute something accurately, economically, and fast on the particular established or selection of computers, making use of development languages and particular resources. This period is really some actions with:

- **Input**
  - Design and user requirements.

- **Output**
  - Expected Result.

- **Methods**
  - Transformation, conversion, monitoring, testing.

The implementation procedure starts with planning a strategy for your execution of the machine. According to this strategy, the actions are to be completed, conversations made regarding the assets and gear and the extra equipment has to be got to apply the new program. In order to carry out above mentioned activities, following points are very important.

- **Functionality**
  - A computational design of the system.

- **Resource**
  - The Computer, languages, tools, and services available to build the system.

- **Performance**
  - The expected response times of the system.

- **Miscellaneous**
  - Quality, scheduling, testing etc.

### 5.6.2 Implementation details
Steps carried out during context based word sense system are described as follows with examples.

Step1: Taking input as XML file which contains text documents. As shown in figure 5.13 applications tested on two different XML file for evaluation. First file contains
information about exchange and second file contains information about cancer with two different contexts.

<table>
<thead>
<tr>
<th>Telephone Exchange</th>
<th>Cancer Diseases and Astrology</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;REUTERS&gt;</td>
<td>&lt;REUTERS&gt;</td>
</tr>
<tr>
<td>&lt;DATE&gt;26-FEB-1987 15:01:01.79&lt;/DATE&gt;</td>
<td>&lt;DATE&gt;26-FEB-1987 15:01:01.79&lt;/DATE&gt;</td>
</tr>
<tr>
<td>&lt;TOPICS&gt;</td>
<td>&lt;TOPICS&gt;</td>
</tr>
<tr>
<td>&lt;D&gt; Telephone Exchange&lt;/D&gt;</td>
<td>&lt;D&gt; Cancer Diseases and Astrology&lt;/D&gt;</td>
</tr>
<tr>
<td>&lt;TEXT&gt;</td>
<td>&lt;TEXT&gt;</td>
</tr>
<tr>
<td>&lt;TITLE&gt; Telephone Exchange&lt;/TITLE&gt;</td>
<td>&lt;TITLE&gt; Cancer Diseases and Astrology&lt;/TITLE&gt;</td>
</tr>
<tr>
<td>&lt;DATELINE&gt;Feb 26 -&lt;/DATELINE&gt;</td>
<td>&lt;DATELINE&gt; Feb 26 -</td>
</tr>
<tr>
<td>&lt;BODY&gt;</td>
<td>&lt;BODY&gt;</td>
</tr>
<tr>
<td>In the area of telecommunications, a telephone exchange or telephone change is just something of digital elements that attatches telephone calls. A main workplace is the bodily building utilized to home inside grow gear including telephone changes, which make phone calls &quot;work&quot; in the feeling of creating contacts and communicating the language info. Reuter</td>
<td>Cancer known as a malignant neoplasm clinically is a phrase for a big number of distinct disorders, all involving unregulated cell growth. In cancer, cells grow and break up uncontrollably, forming malignant tumors, and invade nearby areas of the body. The cancer could also spread to more remote areas of the body through the lymphatic system or bloodstream. Not all growths are malignant. Benign tumors do not grow uncontrollably, do not invade nearby tissues, and do not spread through the body. Reuter</td>
</tr>
<tr>
<td>&lt;/BODY&gt;</td>
<td>&lt;/BODY&gt;</td>
</tr>
<tr>
<td>&lt;/TEXT&gt;</td>
<td>&lt;/TEXT&gt;</td>
</tr>
<tr>
<td>&lt;/REUTERS&gt;</td>
<td>&lt;/REUTERS&gt;</td>
</tr>
</tbody>
</table>
Figure 5.13: XML file with Telephone exchange and cancer document.

Step 2: Application parses above two XML file into text file as follows. While parsing it removes all XML tag and writes text data into text file.

<table>
<thead>
<tr>
<th>Telephone Exchange</th>
<th>Cancer Diseases and Astrology</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the area of telecommunications, a telephone exchange or telephone change is just something of digital elements that attaches telephone calls. A main workplace is the bodily building utilized to home inside grow gear including telephone changes, which make phone calls &quot;work&quot; in the feeling of creating contacts and communicating the...</td>
<td>Cancer identified as a malignant neoplasm clinically is a phrase for a big number of distinct disorders, all involving unregulated cell growth. In cancer, cells grow and break up uncontrollably, forming malignant tumors, and invade nearby areas of the body. The cancer could also spread to more remote areas of the body through the lymphatic system or...</td>
</tr>
</tbody>
</table>
language info. Reuter

bloodstream. Not all growths are malignant. Benign tumors do not grow uncontrollably, do not invade nearby tissues, and do not spread through the body. Reuter

Cancer is the next astrological sign in the Horoscope. It's considered one of four cardinal signs and a water indication. Cancer is dominated by the Moon. People if the Sunlight is within this sign created are considered Cancerian people. Under the tropical zodiac, the Sunshine enters Cancer on the minute of summer by description, or about on June 21, leaving it around July 21. Reuter

**Figure 5.14: Parsed XML file into text file**

Step 3: In this step application uses parsed XML file and common stop words shown in table 5.6 to generate important keywords as shown in fig. 5.15a and 5.15b.

**Figure 5.15a: Generated keyword from Telephone Exchange text file**
Figure 5.15b: Generated keyword from Cancer Diseases and Astrology text file

<table>
<thead>
<tr>
<th>a</th>
<th>below</th>
<th>e</th>
<th>him</th>
<th>near</th>
<th>see</th>
<th>used</th>
</tr>
</thead>
<tbody>
<tr>
<td>able</td>
<td>beside</td>
<td>each</td>
<td>his</td>
<td>nearly</td>
<td>seeing</td>
<td>useful</td>
</tr>
<tr>
<td>about</td>
<td>best</td>
<td>either</td>
<td>i</td>
<td>need</td>
<td>seem</td>
<td>uses</td>
</tr>
<tr>
<td>above</td>
<td>better</td>
<td>else</td>
<td>ie</td>
<td>neither</td>
<td>seems</td>
<td>went</td>
</tr>
<tr>
<td>across</td>
<td>between</td>
<td>elsewhere</td>
<td>if</td>
<td>never</td>
<td>seen</td>
<td>were</td>
</tr>
<tr>
<td>actually</td>
<td>beyond</td>
<td>enough</td>
<td>in</td>
<td>next</td>
<td>she</td>
<td>what</td>
</tr>
<tr>
<td>again</td>
<td>both</td>
<td>etc</td>
<td>indeed</td>
<td>no</td>
<td>should</td>
<td>when</td>
</tr>
<tr>
<td>against</td>
<td>but</td>
<td>even</td>
<td>indicate</td>
<td>nobody</td>
<td>since</td>
<td>whenever</td>
</tr>
<tr>
<td>all</td>
<td>by</td>
<td>every</td>
<td>into</td>
<td>now</td>
<td>six</td>
<td>where</td>
</tr>
<tr>
<td>allow</td>
<td>c</td>
<td>everybody</td>
<td>is</td>
<td>m</td>
<td>so</td>
<td>whereas</td>
</tr>
<tr>
<td>allows</td>
<td>came</td>
<td>everyone</td>
<td>it</td>
<td>many</td>
<td>some</td>
<td>whereby</td>
</tr>
<tr>
<td>almost</td>
<td>can</td>
<td>everywhere</td>
<td>its</td>
<td>may</td>
<td>somehow</td>
<td>which</td>
</tr>
<tr>
<td>along</td>
<td>cannot</td>
<td>f</td>
<td>itself</td>
<td>me</td>
<td>taken</td>
<td>while</td>
</tr>
<tr>
<td>already</td>
<td>cant</td>
<td>far</td>
<td>j</td>
<td>mean</td>
<td>tell</td>
<td>who</td>
</tr>
<tr>
<td>also</td>
<td>cause</td>
<td>few</td>
<td>just</td>
<td>might</td>
<td>t</td>
<td>whoever</td>
</tr>
<tr>
<td>always</td>
<td>certain</td>
<td>following</td>
<td>k</td>
<td>more</td>
<td>than</td>
<td>whom</td>
</tr>
<tr>
<td>among</td>
<td>certainly</td>
<td>follows</td>
<td>keep</td>
<td>must</td>
<td>thank</td>
<td>whose</td>
</tr>
<tr>
<td>an</td>
<td>consider</td>
<td>for</td>
<td>keeps</td>
<td>my</td>
<td>that</td>
<td>why</td>
</tr>
<tr>
<td>and</td>
<td>contain</td>
<td>g</td>
<td>kept</td>
<td>myself</td>
<td>the</td>
<td>z</td>
</tr>
<tr>
<td>another</td>
<td>could</td>
<td>get</td>
<td>know</td>
<td>o</td>
<td>then</td>
<td>zero</td>
</tr>
<tr>
<td>any</td>
<td>currently</td>
<td>given</td>
<td>knows</td>
<td>of</td>
<td>there</td>
<td></td>
</tr>
<tr>
<td>anybody</td>
<td>d</td>
<td>gives</td>
<td>known</td>
<td>off</td>
<td>truly</td>
<td></td>
</tr>
<tr>
<td>anyhow</td>
<td>definitely</td>
<td>go</td>
<td>l</td>
<td>often</td>
<td>try</td>
<td></td>
</tr>
<tr>
<td>anyone</td>
<td>did</td>
<td>h</td>
<td>last</td>
<td>out</td>
<td>trying</td>
<td></td>
</tr>
<tr>
<td>are</td>
<td>different</td>
<td>had</td>
<td>less</td>
<td>over</td>
<td>u</td>
<td></td>
</tr>
<tr>
<td>as</td>
<td>do</td>
<td>has</td>
<td>test</td>
<td>own</td>
<td>un</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>does</td>
<td>have</td>
<td>let</td>
<td>p</td>
<td>until</td>
<td></td>
</tr>
<tr>
<td>be</td>
<td>doing</td>
<td>he</td>
<td>like</td>
<td>please</td>
<td>up</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.6: Common stop word list

<table>
<thead>
<tr>
<th>because</th>
<th>done</th>
<th>her</th>
<th>look</th>
<th>plus</th>
</tr>
</thead>
<tbody>
<tr>
<td>become</td>
<td>down</td>
<td>here</td>
<td>ltd</td>
<td>possible</td>
</tr>
</tbody>
</table>

Step 4: In this step user will input his/her query in the form of word for sensing its meaning with respect to keyword generated in previous step. In following example system evaluate word sense for exchange and cancer. In the context of exchange WordNet has total 11 different senses in its dictionary, but system will output sense with respect to telephone exchange context. It ignores other dictionary meanings related to stock exchange, give and receive contexts. For this it uses Neighborhood Keywords of exchange word for mapping its relation with dictionary synonyms of exchange. Fig. 5.16a shows Generated keyword from Telephone Exchange and its Neighborhood Keywords as follows in which underlined word match relation with WordNet exchange synonyms and display its meaning.

“field telecommunications telephone exchange telephone switch system electronic components connects telephone calls central”

Similar process is carried out for word cancer with respect to Cancer Diseases and Astrology document. Here cancer word associated with two different contexts namely Diseases and Astrology. System will display two different senses with respect to these two contexts as shown in fig. 5.16b.

![Figure 5.16a: User word sense result for context exchange](image-url)
5.7 SOFTWARE TESTING AND QUALITY ASSURANCE

Computer software testing is an analysis conducted to provide stakeholders with details concerning the caliber of the merchandise or service under evaluation. Software testing may also offer an objective, separate perspective of the software to enable the company to comprehend and recognize the dangers of software execution. Evaluation methods comprise, but aren't restricted to, the procedure of running a software or program with the aim of discovering software bugs (errors or other problems).

- Fulfills the conditions that guided its development and design development
- Functions as expected
- Could be applied with the same features
- Fulfills the requirements of stakeholders.

Software testing, based on the screening procedure applied, may be applied anytime within the development procedure. Typically most of the check work happens after the conditions have been described and the code procedure has been finished, however in the Agile techniques most of the test effort is on-going. As a result, the strategy of the evaluation is ruled by the preferred computer software improvement strategy. The test effort will be focused by different software development models at different factors within the development process. Test is often employed by newer development models, such as Agile, - driven place and development an elevated part...
of the screening in the fingers of the programmer, before it reaches a proper group of evaluators. In the code procedure has been finished and a far more conventional design, the majority of the check performance happens following the conditions have been described.

Quality assurance (QA) relates to the systematic activities in a quality program so that quality needs for an item or service may be satisfied. It's error prevention that is conferred by an associated feedback loop, observation of procedures and the organized measurement, assessment using a regular. This is often compared with quality handle, that is centered on procedure results. Two concepts contained in QA are: "Fit for purpose", the merchandise should be acceptable for the intended purpose; and "Correct first-time", blunders should be removed. QA includes administration of the caliber of products, assemblies, raw materials and parts, providers related to production, and production, administration and review procedures. Appropriate quality is decided by item customers, clients or customers, maybe not by culture generally. It's not associated with price and adjectives or descriptors such "high" and "inferior" are perhaps not appropriate. For instance, a inexpensive merchandise might be looked at as having large quality because it is throw-away where another may be viewed as having bad quality because it isn't throw-away.

5.7.1 Software Testing

Software screening is just a confirmation procedure for advancement and software quality evaluation. The chief aim of screening computer software item is to discover insects in applications before providing to consumer. Still some bugs remain in software product due to incomplete testing. Even with this incomplete testing process, the importance of testing should not be underestimated. Software testing determines software quality assurance and signifies the evaluation of specification, layout, and code era. An important role is played by testing in evaluating and attaining the standard of software product.

5.7.2 Objective of Software Testing

i) The goal of testing is to find defects before customers find them out.

ii) The objective of testing is to show that the system works, which gain developer team confidence in unit testing.
iii) Most of the complex software systems contain faults, which cause the system to fail from time to time. As faults are discovered and fixed while performing more and more tests, the failure rate of a system generally decreases.

iv) The highest level of objective of performing tests is to produce low-risk software with less number of test cases. This idea leads us to the concept of effectiveness of test cases.

5.7.3 What’s a Test Case?

In its most essential type, a check circumstance is really an easy set of. In systems, where the result depends entirely about the present input signal, test cases are quite basic in arrangement. A compiler for the D programming vocabulary is instance of the stateless program. A compiler is a stateless method because to gather an application it doesn't have to understand about the applications it gathered previously. In state-oriented systems, where the plan result depends both about the current condition of the current input signal and also the machine, a test situation might comprise of a series of sets.

5.7.4 Software Testing Types

5.7.4.1 Unit Testing

Unit testing describes testing software models in seclusion. A few examples of generally recognized models are functions, processes, or approaches. Being a software component actually a course within an object-oriented programming terminology can be viewed. The perform done by a course device might not need an immediate connection with a system-level perform. Therefore, a plan device might be looked at like a bit of signal applying a "low" degree perform. You will find two good reasons for screening a device in a stand-alone fashion. First, problems discovered throughout screening can be credited to some particular device such that it can be readily repaired. Furthermore, unit testing removes dependencies on additional software models. Second, during device testing it's desired to check that each unique delivery of the software unit creates the anticipated consequence. In phrases of signal particulars, a distinct execution describes a distinct route within the system. Preferably, all potential or around feasible distinct accomplishments should be looked at during device testing. This demands cautious choice of enter information for every delivery. A developer has immediate use of the input signal vector of the system by
running a plan unit in seclusion. This immediate entry makes it simpler to run as several distinctive pathways as appealing or feasible.

Unit screening is conducted by the program unit is written by the programmer who because the programmer is closely acquainted with the inner information on the device. The aim for the developer is to become happy that the system functions as expected. A unit check is done by him, because a developer is designed to assemble a device with no mistakes inside or her for their satisfaction in the starting and also to the fulfillment of additional developers when the unit is incorporated with other models. In this system the testing is started by unit testing in that each module and sub modules are tested. Following example in this system tested for individual program unit.

Example:
For creating keyword file, initially it takes each token from text file in str2. Then str2is compared for numeric value. If it is numeric it is skipped otherwise it is written in keyword file.

```
if(isNumeric(str2))
{
    Continue; // If true Skip current str2 token.
}
public boolean isNumeric(String s)
{
    return s.matches("((-|\+)?[0-9]+(\.[0-9]+)?)+");
}
```

**Figure 5.17: Sample Program unit for preprocessing**

5.7.4.2 Integration Testing
Integration screening is especially targeted at subjecting the difficulties that appear in the blend of parts. In integration testing smaller models are incorporated into larger units and larger units into the complete program. This varies from device screening in that units are not any longer analyzed alone but in organizations, the emphasis changing from the personal units for the conversation between them. They could be incorporated and assessed to the complete job as effective places of performance are finished. Without integration testing you're restricted to testing a fully constructed
item or program that is ineffective and problem prone. Integration assessments are made to analyze built-in computer software parts as you application when they really operate to ascertain.

This system has two main modules as preprocessing and word sense. These two modules first tested by unit testing then it tested for integration into application.

5.7.4.3 Functional Testing
Three crucial aspects of a perform are its input, result, and anticipated change of input to result. Disregarding the facts of the genuine change of feedback to result, it assesses the domains of the input and also the result factors of applications to create check information. Practical screening is dedicated to the next items:

Valid Input: recognized courses of valid input should be recognized.
Invalid Input: recognized courses of invalid input should be declined.
Functions: recognized functions should be practiced.
Result: recognized courses of program results should be practiced.
Systems/Procedures: interfacing methods or methods should be invoked.

Business and planning of practical assessments is concentrated on demands, essential capabilities or particular test circumstances. In addition, methodical coverage relating to determine Enterprise process runs; data areas, predefined processes, and consecutive procedures should be looked at for screening. Before practical screening is total, extra tests are recognized and the efficient value of present tests is decided.

The system is checked for various inputs for validation. In following example of code unit check acceptance of input file.

Example:
For checking valid input XML file.

```java
if(txtXMLLable.getText() != null && txtXMLLable.getText().contains(".xml"))
{
    System.out.println("Correct xml file");
    getData(txtXMLLable.getText());
    textArea.setText(finalKeywordStr);
    JScrollPane1 = new JScrollPane(textArea);
    addComponent(cpane, JScrollPane1, 10, 120, 880, 200);
    setDefaultCloseOperation(HIDE_ON_CLOSE);
    setVisible(true);
}
```
else

{ JOptionPane.showMessageDialog(new JFrame(), "Please Enter valid XML file or browse it", "Dialog", JOptionPane.ERROR_MESSAGE);

}

Figure 5.18: Sample Program unit for checking valid input

5.7.4.4 Black Box Testing

Testing of finished models of practical signal is well known as black-box screening because testers handle like a black the item-box. They concern themselves with confirming given input signal against anticipated output signal rather than fretting about the reasoning of what goes on between. User Acceptance Testing (UAT) and Methods Testing are classic instance of black-box testing. This is a screening where the program under check is handled, like a black container, you can't "see" in to it. The evaluation gives advices and reacts to results without contemplating the way the computer software functions. In this system there are many functionalities have been checked for black box testing such as acceptance of file, displaying preprocessing results, accepting word to be sensed, displaying word sense result. In this end user need not to know what functionality is used inside these modules?

5.7.4.5 White-box Testing

Every software product is realized by means of a program code. White box testing is really a manner of testing the exterior performance of the signal by testing and analyzing the application code that recognizes the external performance. Available box testing or this really can also be referred to as clear box or glass box. White-box or glass-box testing depends on examining the signal itself and the inner reasoning of the applications. It utilizes methods including extremely specialized or technologies specific screening through issues like signal inspections. Although white-box methods may be used at any period in a computer software product's life period they are usually present in Device testing actions.

In this system code is inspected for working of different functionality as under:

- Parsing XML file.
- Extracting Keywords.
• Sensing word.
• Displaying results.

5.7.4.6 Regression testing
Regression testing is done each time a part of the machine is altered. The crucial notion in regression testing is always to assess that the customization has not launched any fresh problems in the part that was not susceptible to customization. To be exact, regression testing isn't a different amount of testing. Instead, it's regarded as a sub period of device, integration, and system-level testing. In regression testing, fresh assessments aren't created. Rather, assessments are chosen, prioritized, and performed from the present pool of check instances to ensure nothing is damaged in the fresh variant of the applications. Regression testing can be a costly procedure and makes up about a prevalent part of testing work within the business. It is desired to choose a part of the test instances from the present swimming to lessen the price. A crucial query is how many and so the selected test cases may discover new problems which test instances must be selected. Changes were made in the code at each stage. For testing and correction in functionality performance regression testing method is used.

5.7.4.7 GUI Graphical User Interface Testing
GUI (Graphical Interface) testing would be the testing recommendations for specific surroundings, architectures and generally experienced information program by computer software technicians. Due to multiple-use parts supplied being part of GUI development conditions, the generation of interface is becoming less time intensive and more exact. As modern GUIs have the same look and feel, a series of standard tests can be derived. In this system main GUI checked for three window screens. GUI is tested for main menu, Preprocessing and word sense window. Fig. 5.19 shows GUI testing for acceptance of file.
5.7.5 Sample Test Case and Test Results

<table>
<thead>
<tr>
<th>Test ID</th>
<th>Test Case Data</th>
<th>User Action</th>
<th>Expected Result</th>
<th>Actual Result</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>TID1</td>
<td>option &quot;Preprocessing&quot;</td>
<td>Select preprocesses XML file</td>
<td>&quot;Preprocessing page is displayed&quot;</td>
<td>&quot;Preprocessing page is displayed&quot;</td>
<td>working properly</td>
</tr>
<tr>
<td>TID2</td>
<td>option &quot;Word Sense&quot;</td>
<td>selected word sense in uploaded file</td>
<td>&quot;Word Sense page is displayed&quot;</td>
<td>&quot;Word Sense page is displayed&quot;</td>
<td>working properly</td>
</tr>
<tr>
<td>TID3</td>
<td>User selects preprocessing</td>
<td>upload XML file</td>
<td>&quot;XML file loaded successfully&quot;</td>
<td>&quot;XML file loaded without any error&quot;</td>
<td>working properly</td>
</tr>
<tr>
<td>TID4</td>
<td>upload file other than XML</td>
<td>User select text file and click OK</td>
<td>Message displayed &quot;please select valid file&quot;</td>
<td>Message prompted &quot;please select valid file&quot;</td>
<td>working properly</td>
</tr>
<tr>
<td>TID5</td>
<td>User press OK for preprocessing XML file</td>
<td></td>
<td>&quot;Keyword file generated and displayed on screen&quot;</td>
<td>&quot;Keyword file displayed on screen&quot;</td>
<td>working properly</td>
</tr>
<tr>
<td>TID6</td>
<td>Check token for punctuation character</td>
<td>User press OK for preprocessing XML file</td>
<td>Punctuation token should be ignored</td>
<td></td>
<td>working properly</td>
</tr>
</tbody>
</table>
Table 5.7: Sample Test Case and Test Results

5.7.6 Quality Assurance

The primary aim of a quality assurance strategy would be to develop standards for the approval of the results and to identify the predicted results and the targets. Its intention is to explain the guidelines and the processes to become employed and the resources to become utilized so that you can ensure the quality all through the development, layout, testing and approval stages of the undertaking. The quality assurance task for the work-done is mainly performed at three different levels: a) Requirements b) Design Level c) Development Level. Now, according to quality achievement plan prepared earlier, tasks such as traceability matrix, quality checklist and criticality analysis is performed in each of the levels as and when required. Requirements Phase:

This ensures that the software requirements are:

- Complete
- Testable
- Properly expressed as functional, performance and interface requirements.

Design Phase:

- Ensure that approved design standards are followed
- Ensure that allocated modules are included in the detailed design
- Ensure that the results of design inspections are included in the design
- Review the Critical Design Documentation and make sure that all action items are resolved.

Implementation Phase:
• Results of coding and design activities including the schedule contained in the Software Development Plan
• Status of all deliverable items
• Configuration management activities
• Nonconformance reporting and corrective action system

5.7.6.1 Quality assurance at Requirement Analysis Phase

Requirement Analysis provides following documents.

i) Functional Specification
ii) Non-Functional Specification

i) Functional Specification:

Functional Specification is the very first phase of Software Development Life Cycle. High quality assurance much more depends on this phase because it gives you defects, traceability, verification and validation in the early phase of system development. Even if implementation is not started still you can review each module of the system by preparing quality assurance checklist document. Both development and quality team involved at this phase to perform quality assurance. For application purpose following functional specification is prepared.

a) Pre-requisites:

Before using context based word sense extraction in text application require following resources.

• Installed WordNet Dictionary of minimum version 2.1 for sensing a word.
• Reuter 21578 Dataset and its extracted important keywords.
• Installed Java Environment for running system successfully.

b) Input Requirements:

• Reuter 21578 XML Dataset.
• Reuter 21578 Parsed XML Files.
• Reuter 21578 Extracted Keywords.
• Word entered by user sense its meaning.

c) Output Requirements

• Reuter 21578 Parsed XML Files.
• Reuter 21578 Extracted Keywords.
• Word sense according to context.
d) Workflow of system

- System will take Reuter 21578 XML files and parse it into new text files.
- Form parsed text file system will extract important keywords of Reuter 21578 documents.
- User will input ‘word’ to be sensed in Reuter 21578 documents.
- System will search Reuter 21578 documents for matching user input ‘word’.
- For matched Reuter 21578 documents, system will collect neighbourhood keyword of inputted ‘word’ in temporary buffer.
- System will extract all synonyms of inputted ‘word’ from dictionary and compare it with stored neighborhood keyword of inputted ‘word’.
- If match found then it will display corresponding sense of inputted ‘word’.

ii) Non-Functional Specification:

a) Developmental Constraints:
The system will depend on the Reuter 21578 Dataset and its extracted important keywords. It uses interface WordNet dictionary to sense word.

b) Dataset Constraint:
In this work specific Reuter 21578 Dataset used for performing word sense on text data. The system is limited for dataset into xml format. Another data used in this work is WordNet dictionary which gives word sense.

c) Time Constraint:
The system is confined to be completed in specified period. A schedule will work out which made it even more challenging. For inspecting quality assurance this document is used as baseline. The quality assurance is performed while working on above phase by reviewing each requirement with strictly checkmarks. Quality assurance checks functionality of system and added addressed suggestions to minimize defect or clarification against the Functional Specification document. Checklist is prepared for application which provides a complete list of items to be verified and also provide space for document findings of the checks performed. Table 5.8 shows reviews of requirement specification documents.
I. Checklist for Requirement Specification

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Topic/Requirement</th>
<th>Yes</th>
<th>No</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>The Document</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>Is the document prepared according standard template?</td>
<td>Yes</td>
<td></td>
<td>Standard template like <em>Software Requirement Specification</em> is used to collect requirement.</td>
</tr>
<tr>
<td>b</td>
<td>Is prepared document complete?</td>
<td>Yes</td>
<td></td>
<td>Prepared Document Inspected and found in completed manner.</td>
</tr>
<tr>
<td></td>
<td><strong>Functional Specification</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>Required functions were properly and clearly defined?</td>
<td>Yes</td>
<td></td>
<td>All main function in system is decided.</td>
</tr>
<tr>
<td>d</td>
<td>Designed inputs conform to required outputs?</td>
<td>Yes</td>
<td></td>
<td>Manually input work traced to expected results.</td>
</tr>
<tr>
<td>e</td>
<td>Is GUI interfaces are fully defined?</td>
<td>Yes</td>
<td></td>
<td>Core part of GUI is decided.</td>
</tr>
<tr>
<td>f</td>
<td>Is response time, input flow capacity, storage capacity are correctly defined?</td>
<td></td>
<td>No</td>
<td>Rough calculation is done for storage and response time but actual result be found as system progress is done.</td>
</tr>
<tr>
<td>g</td>
<td>Is WordNet dictionary interface with Context Based Word Sense Extraction in Text are correctly defined?</td>
<td>Yes</td>
<td></td>
<td>WordNet dictionary interface with system is done.</td>
</tr>
<tr>
<td>h</td>
<td>Step designed to test fulfillment of the specified requirement are correctly working?</td>
<td>Yes</td>
<td></td>
<td>Each module tested according to workflow.</td>
</tr>
<tr>
<td>i</td>
<td>Is Plan ready for handling</td>
<td>Yes</td>
<td></td>
<td>System work is research based</td>
</tr>
</tbody>
</table>
change requests during development? | so changes are handled by reverse engineering.

### III Non-Functional Specification

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Are the specified requirements feasible considering the available Resources and budget?</td>
<td>Yes</td>
</tr>
<tr>
<td>b</td>
<td>Can system completed in predefined schedule?</td>
<td>Yes</td>
</tr>
<tr>
<td>c</td>
<td>Is risks has been identified?</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Remarks:** System is achieving quality in requirement analysis as it performs positive checklist.

#### Table 5.8: Reviews of Requirement Specification Documents

From above review for word sense extraction based on its context in text it shows following quality assurance points.

- Checklists support document completeness and improve document quality as all the relevant items and quality issues to be reviewed are already listed.
- Conduct of review sessions becomes less problematic when topics and their order of priority are defined and well known. An efficient session is expected to carry out a thorough analysis of comments by reviewers.

## II. Criticality Analysis:

Criticality analysis includes various critical issues at requirements level, if not taken care of, impact on the design phase. Critical issues to be taken care of at requirements & design level are:

- Requirements disambiguation.
- Clear understanding of problem statement.
- Clear understanding of abstract view of system.
5.7.6.2 Quality assurance at Design phase

A really significant stage in computer software development life period after software requirement evaluation is a layout stage, in which the advancement of the development function done is documented in the kind of design record. During the design phase, the logical thinking described in the system requirements specification document, is systematically transformed into the physical representation of the system. In other words, the requirements tell the developer “what to do” while the design tells the developer “how to do it.” The primary verification activity during the design phase is the formal review of the preliminary and detailed design documents. These documents are verified for their consistency, completeness, and correctness within themselves and with the requirements specification document. The software design is evaluated to ensure that all the functionality of the requirements specification document is present and that no extraneous functionality is present. A deficiency of quality within the look procedure can invalidate good specifications standards and can make right execution difficult. Used it implies that use of record throughout design aids enhance design quality. The design document is prepared which is checked repeatedly, in order to detect any possible error that might have entered. This will examine design document thoroughly before granting their approval for development. For design phase in research work entitled ‘Word Sense Extraction based on its Context in Text’ following design review prepared by checklist to achieve Quality assurance.

I. Design Review by checklist:

Following checklist captures elements that are required for system design. It's prepared throughout the Design Evaluation procedure to make sure the design being defined features all appropriate design factors. Following design gives Traceability to Requirements.
## Checklist for Design Review

**Title:** A Novel Design Approach For Word Sense Extraction Based on Its Context

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Topic/Requirement</th>
<th>Yes</th>
<th>No</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I</strong></td>
<td><strong>Requirement Traceability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>Does the design address all issues specified in functional requirements?</td>
<td>Yes</td>
<td></td>
<td>All the basic functionality is prepared according to requirements.</td>
</tr>
<tr>
<td>b</td>
<td>Is design traceable back to requirements i.e. Additional new feature is adjustable with earlier phase?</td>
<td>Yes</td>
<td></td>
<td>New function such as keyword finding is adjusted with existing features.</td>
</tr>
<tr>
<td>c</td>
<td>Are the specifications and requirements complete?</td>
<td></td>
<td>No</td>
<td>Some additional function is required like parsing of reuter 21578 documents.</td>
</tr>
<tr>
<td>d</td>
<td>Is all requirement documents traced with a traceability matrix?</td>
<td>Yes</td>
<td></td>
<td>Traceability matrix is generated from requirements.</td>
</tr>
<tr>
<td><strong>II</strong></td>
<td><strong>Functionality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>Are decision making correctly stated in the workflow logic?</td>
<td>Yes</td>
<td></td>
<td>Number decision in system is specified according to workflow.</td>
</tr>
<tr>
<td>b</td>
<td>Does error or exception properly handled?</td>
<td>Yes</td>
<td></td>
<td>Exceptions are caught and throw out.</td>
</tr>
<tr>
<td>c</td>
<td>Does it fulfill to designed architecture?</td>
<td>Yes</td>
<td></td>
<td>Works according requirement described in system architecture.</td>
</tr>
<tr>
<td>d</td>
<td>Will the design be easy to port to another environment if appropriate?</td>
<td>No</td>
<td></td>
<td>Currently testing on platform specified in requirement.</td>
</tr>
<tr>
<td>e</td>
<td>Does information flow, control flow and interfaces are clearly symbolized?</td>
<td>Yes</td>
<td></td>
<td>Specified by dataflow diagrams, flowcharts and UML diagrams etc.</td>
</tr>
</tbody>
</table>
### III Design Centered

<table>
<thead>
<tr>
<th></th>
<th>Design Centered</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Does the design support both product and system goals?</td>
<td>Yes</td>
<td>System goal creates final product.</td>
</tr>
<tr>
<td>b</td>
<td>Does the design grant approval to the next development step?</td>
<td>Yes</td>
<td>Complete design will permit to start development of product.</td>
</tr>
<tr>
<td>c</td>
<td>Does the design follow standard techniques to describe system?</td>
<td>Yes</td>
<td>Designed according to requirement and uses standard diagrams like UML, DFD etc.</td>
</tr>
<tr>
<td>d</td>
<td>Is the interface consistent with the design?</td>
<td>Yes</td>
<td>Use case describes user interaction with system.</td>
</tr>
</tbody>
</table>

### IV Design Feasibility and Reliability

<table>
<thead>
<tr>
<th></th>
<th>Design Feasibility and Reliability</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Is the design feasible from a technology, cost, and schedule standpoint?</td>
<td>Yes</td>
<td>Cost is affordable and system completed within time.</td>
</tr>
<tr>
<td>b</td>
<td>Are boundary checks performed on memory accesses to ensure program memory overflow or underflow?</td>
<td>Yes</td>
<td>Text file checked for 1000 documents.</td>
</tr>
<tr>
<td>c</td>
<td>Are undesired events considered? Like restart Functions.</td>
<td>No</td>
<td>Tracing will be done as work progress.</td>
</tr>
</tbody>
</table>

**Remarks:** By above checklist of design review, it gives satisfaction to application requirement traceability and permit further development stage.

*Table 5.9: Reviews of Design Document*
II. Traceability Matrix

Traceability matrix at this level is prepared by taking requirements of the project as columns and design modules as rows.

<table>
<thead>
<tr>
<th></th>
<th>Parse Text from Reuter File</th>
<th>Parse Keyword from text file</th>
<th>Word Sense input</th>
<th>Display Word Sense</th>
<th>Module Communication</th>
<th>V&amp;V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Design for Reuter file</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Design Input Text File</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Input Design for Word Sense</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Design of Display Sense</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Design for Modules-Communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

Table 5.10: Traceability matrix at design level

III. Criticality Analysis

Criticality analysis includes various critical issues at design level, if not taken care of, impact on the next phase. Critical issues to be taken care of at design level are:

- Clear understanding of functional requirements and non-functional requirements
- Clear understanding design modules as per requirements

5.7.7 Quality assurance at Implementation phase

Various kinds of Active Layout Evaluations - Reviews, Code - Inspections, Code and Walkthroughs have been discovered to be quite helpful in enhancing the quality of
applications. Signal assessments should confirm that it tools the prepared layout, test for technical correctness and completeness of the signal, and ensure good programming methods and requirements are utilized. Checklists for code-reviews are generally said in query forms focusing on tips, allowing the recognition of a software problem or of a course of problems. Throughout as helpful information for examination of signal an evaluation of line by line application, record can be used. Detected problems are categorized and eliminated according to their relevance after the conclusion of the evaluation, and the first record is enhanced according for the lately detected problems. The implementation document is prepared which is checked repeatedly, in order to detect any possible error that might have entered. Record function would be to make sure that process attain the greatest requirements comparable to demonstration, uniformity, completeness, quality, and structure. For implementation phase following Implementation Review is prepared by checklist to achieve Quality assurance.

I. Implementation Review by checklist.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Topic/Requirement</th>
<th>Yes</th>
<th>No</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Interfaces ,Code and data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>Are Interface done as specified by the design?</td>
<td>Yes</td>
<td></td>
<td>User interaction for word sense is developed.</td>
</tr>
<tr>
<td>b</td>
<td>Does the code track the Detailed Design?</td>
<td>Yes</td>
<td></td>
<td>Basic functionality of each module remains same as specified by design.</td>
</tr>
<tr>
<td>c</td>
<td>Are values returned from functions calls checked before proceeding to the following code?</td>
<td>Yes</td>
<td></td>
<td>Values like parsed xml file, generated keywords checked which is stored in files.</td>
</tr>
<tr>
<td>d</td>
<td>Are the variables initialized with predefined value?</td>
<td>Yes</td>
<td></td>
<td>Flag setting used for matching dictionary synonym and neighborhood keyword.</td>
</tr>
</tbody>
</table>
e | Should this system be doing correct I/O? | Yes | It takes input as single word and display sense of that word.
---|---|---|---
f | Are decision making handled correctly? | Yes | Matching word in reuter and dictionary gives correct decision idea.

**II Feasibility and Reliability**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
a | Is the system feasible from a technology, cost, and schedule standpoint? | Yes | Cost is affordable and system development is completed within time.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
b | Is there backup copies maintained for developed software? | Backup copies of the developed software have been stored in a desktop as well as on disc.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
c | Is system is feasible to make extension? | Yes | Currently work on only word sense but in feature development possible for all text mining module.

Remarks: Within the assessment procedure of Execution period for the Standard Guarantee Strategy this listing is provided.

**Table 5.11 Reviews of Implementation Document**

II. Traceability Matrix

Traceability matrix at this level is prepared by taking design modules as columns and code modules of the project as rows.

<table>
<thead>
<tr>
<th></th>
<th>Textbox (for input)</th>
<th>Design of Create</th>
<th>Design of Word Sense algorithm</th>
<th>Display Sense</th>
<th>Design for Modules Communication</th>
<th>V&amp;V</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUI code containing ‘Textbox’ and ‘Search’ button</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>
Code for Create textfile | * |  |  | Yes
--- | --- | --- | --- | ---
Code for create keywords documents | * |  |  | Yes
Code For Word Sense | * |  |  | Yes
Code for Modules-Communication |  |  |  | No

Table 5.12: Traceability matrix at Implementation level

III. Criticality Analysis

Criticality analysis includes various critical issues at development level, if not taken care of, impact on the system working. Criticality analysis at this level also includes time complexity of the two modules namely, Word Sense and Tokenization.

Critical issues to be taken care of at development level are:

- Clear and unambiguous code modules
- Functionally correct code modules
- Proper GUI coding
- Unambiguous operations at GUI level

5.8 RESULTS AND DISCUSSION

Among the studies reported all through the preceding study, it is apparent that it is quite hard to evaluate one pair of hence, and effects one approach, with another. The insufficient assessment results from considerable variations in check problems from research to examine. For example, distinct kinds of texts are required, including both extremely specialized or domain name specific texts where sense use is restricted. The outcome dialogue for circumstance established word feeling in text is indicated in terms of Remember and Accuracy. Remember and accuracy terms determines quality of text access by program. Accuracy requires all gathered files under consideration, but the topmost results are considered only by it came back from the machine. For instance for a text search on the set of files accuracy is the amount of right results split by the number of all came back results. Remember in details access is the portion of the files that are applicable for the issue that are effectively recovered. For instance
for text research on the group of records remember is the amount of right results split by the number of results which should have already been returned.

Table 5.13 shows results of precision and recall value for context based word sense in text evaluation.

<table>
<thead>
<tr>
<th>Document Name</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reut2-001</td>
<td>61</td>
<td>64</td>
</tr>
<tr>
<td>Reut-020</td>
<td>60</td>
<td>65</td>
</tr>
<tr>
<td>Doc1</td>
<td>66</td>
<td>69</td>
</tr>
<tr>
<td>Doc2</td>
<td>60</td>
<td>63</td>
</tr>
<tr>
<td>Doc3</td>
<td>61</td>
<td>64</td>
</tr>
</tbody>
</table>

Table 5.13: Precision and Recall value for context based word sense

Figure 5.20 shows the graph plotted for precision and recall results for context based word sense in text.

Figure 5.20: Precision and Recall results for context based word sense in text

Desk 5.14 shows Reuter 21578 Document group features with WordNet for feeling word in Reuter 21578 Document. Here table exhibits qualities for two classes with number of number phrases and files in that files. Last line displays percent of feeling (noun) for comparable Reuter file phrases identified in WordNet book.
Table 5.14: Reuter 21578 Document collection characteristics with WordNet

<table>
<thead>
<tr>
<th>Reuter Document</th>
<th>No. of Documents</th>
<th>Average Length</th>
<th>WordNet Nouns (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn</td>
<td>168</td>
<td>99</td>
<td>60</td>
</tr>
<tr>
<td>Wheat</td>
<td>221</td>
<td>83</td>
<td>62</td>
</tr>
</tbody>
</table>

The Reuter Document column explains the source of the related dataset, the No. of Documents column describes the quantity of documents designated to each class, the Average Span column describes the typical number of phrases in each document and the WordNet Nouns column describes the percent of words in the corresponding documents which are Term Web nouns.