CHAPTER 1

INTRODUCTION

1. INTRODUCTION

This chapter discuss about the problem statement, objectives and features of Auto Text Summarization application.

1.1 Introduction

A summary can be defined as a text that is produced from one or more texts, that contain a significant portion of the information in the original text, and that is no longer than half of the original text.

It is the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user (or users) and task (or tasks).Summaries are an important tool for familiarizing oneself with a subject area.

Text summaries are essential when forming an opinion on if reading a document in whole is necessary for our further knowledge acquiring or not. In other words, summaries save time in our daily work. To write a summary of a text is a non-trivial process where one has to extract the most central information from the original text, and at the same time has to consider the reader of the text and her previous knowledge and possible special interests.

What exactly makes a summary beneficial is an elusive property. Generally speaking there are at least two properties of the summary that must be measured when evaluating summaries and summarization systems –

- Compression Ratio, i.e. how much shorter the summary is than the original,
- Retention Ratio, i.e. how much of the central information is retained.

This can for example be accomplished by comparison with existing summaries for the given text. One must also evaluate the qualitative properties of the summaries, for example how coherent and readable the text is. This is usually done by using a panel of human judges. Furthermore, one can also perform task-based evaluations where one tries to discern to what degree the resulting summaries are beneficent for the completion of a specific task.

Evaluating summaries and automatic text summarization systems is not a straightforward process. Summarization is a complex task that requires understanding of the document content to determine the importance of the text. Lexical cohesion is a method to identify connected portions of the text based on the relations between the words in the text. Lexical cohesive relations can be represented using lexical chains. Lexical chains are sequences of semantically related words spread over the entire text. Lexical chains are used in variety of Natural Language Processing

(NLP) and Information Retrieval (IR) applications.
In current report, we propose a lexical chaining method that includes the glossary relations in the chaining process. These relations enable us to identify topically related concepts, for instance dormitory and student, and thereby enhances the identification of cohesive ties in the text. We then present methods that use the lexical chains to generate summaries by extracting sentences from the document.

1.2 Problem Statement

Popularity of the internet has contributed towards the explosive growth of online information. Search engines provide a means to access huge volumes of information by retrieving the documents considered relevant to the user's query. Even with search engines, the user has to go through the entire document content to judge its relevance. This contributes towards a well-recognized information overload problem.

Similar information overload problems are also faced by corporate networks, which have information spread across various kinds of sources - documents, web pages, mails, faxes, manuals etc. It has become a necessity to have tools that can digest the information present across various sources and provide the user with condensed form of the most relevant information. Summarization is one such technology that can satisfy these needs.

Summaries are frequently used in our daily life to serve variety of purposes. Headlines of news articles, market reports, movie previews, abstracts of journal articles, TV listings, are some of the commonly used forms of summaries.

1.3 Objectives

Today there are numerous documents, papers, reports and articles available in digital form, but most of them lack summaries. The information in them is often too abundant for it to be possible to manually search, shift and choose which knowledge one should acquire. This information must instead be automatically filtered and extracted in order to avoid drowning in it.

Need for abstracts or summary:

- Abstracts promote current awareness
- Abstracts save reading time
- Abstracts facilitate selection
- Abstracts facilitate literature searches
- Abstracts improve indexing efficiency
- Abstracts aid in the preparation of reviews

Automatic Text Summarization is a technique where a computer summarizes a text. A text is given to the computer and the computer returns a shorter less redundant extract of the original text. So far automatic text summarization has not yet reached the quality possible with manual summarization, where a human interprets the text and writes a completely new shorter text with new lexical and syntactic choices. However, automatic text summarization is untiring, consistent and always available.

Document summarization is gaining demand with the explosive growth of online news sources. It requires identification of the several themes present in the document to attain good compression and avoid redundancy. In this report, we propose methods to group the portions of the texts of a document into meaningful clusters. Clustering enable us to extract the various themes of the document collection. Sentences from clusters can then be extracted to generate a summary for the document collection. Clusters can also be used to generate summaries with respect to a given query. We designed a system to compute lexical chains for the given text and use them to extract the salient portions of the document.

CHAPTER 2

LITERATURE SURVEY

2. LITERATURE SURVEY

2.1 Different types of summaries

The various types of summaries which can be generated are

2.1.1 **Generic summaries** aimed at a broad readership community and written by authors or professional abstractors served as surrogates for full-text. However, as our computing environments have continued to accommodate full-text searching, browsing, and personalized information filtering.

2.1.2 **User-focused summaries** have assumed increasing importance. Such summaries rely on a specification of a user information need, such as an area of interest, topic, or query. It should be borne in mind that the notion of a truly generic summary is problematic, since some background assumptions of an-audience are involved in every case.

2.1.3 **Indicative summaries,** which are used to indicate what topics are addressed in the source text, and thus can be used to alert the user as to the source content, and informative summaries, which are intended to cover the concepts in the source text to the extent possible given the compression rate.

2.2 Text Summarization Techniques

There are several ways in which one can characterize different approaches to text summarization are:

2.2.1 **Examine the level of processing**: Based on this, summarization can be characterized as approaching the problem at the surface, entity or discourse levels.

2.2.2 **Surface-level approaches** tend to represent information in terms of shallow features which are then selectively combined together to yield a salience function used to extract information. These features include:

• Thematic features (presence of statistically salient terms, based on term frequencies statistics)

• Location (position in text, position in paragraph, section depth, particular sections)

• Background (presence of terms from the title or headings in the text)

• Cue words and phrases (e.g. in-text summary cues such as “in summary”, “our investigation”, emphasizes such as “important”, “in particular” as well as domain specific ‘bonus’ and ‘stigma’ terms.)

2.2.3 **Entity-level approaches** build an internal representation for text, modeling text entities and their relationships. These approaches tend to represent patterns of connectivity in the text to help determine what is salient. Relationships between entities include:

- Similarity (e.g. Vocabulary overlap)
- Proximity (distance between text units)
- Thesaurus relationships among words
- Co-reference (i.e. of referring expressions such as noun phrases)
- Syntactic relations (based on parse trees)

2.2.4 **Discourse-level approaches** model the global structure of the text, and its relation to communicative goals. This structure can include:

- Format of the document
- Threads of topics as they are revealed in the text
- Rhetorical structure of the text, such as argumentation or narrative structure

2.3 Tools available for summarization:

Various tools available for text summarization are:

2.3.1 **Word Net 2.0 Dictionary:**

Word Net is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, and adjectives are organized into synonym sets, each representing one underlying lexical concept. Different relation link the synonym sets. Such a database may be useful to this project when trying to identify the possible list of "key entities".

2.3.2 **Stanford Natural Language Processor:** The Stanford dependencies provide a representation of grammatical relations between words in a sentence. They have been designed to be easily understood and effectively used by people who want to extract textual relations. Stanford dependencies (SD) are triplets: name of the relation, governor and dependent.

2.4 Approaches

Summarization process is done in following manner:

- **Analysis**: This phase builds an internal representation of the source.
- **Transformation**: This phase generates a representation of the summary based
- **Synthesis**: This phase interprets summary representation back into the natural language form.

Only methods involving multi-document summarization or abstract generation go through the transformation phase. Methods to generate extracts for single document directly go to the synthesis phase after the analysis phase. Each phase undergoes one or more of the following basic condensation operations:

- **Selection**: To filter unimportant and redundant information.
- **Aggregation**: To group information from various portions of the document.
- **Generalization**: To substitute a concept with more general or abstract one.

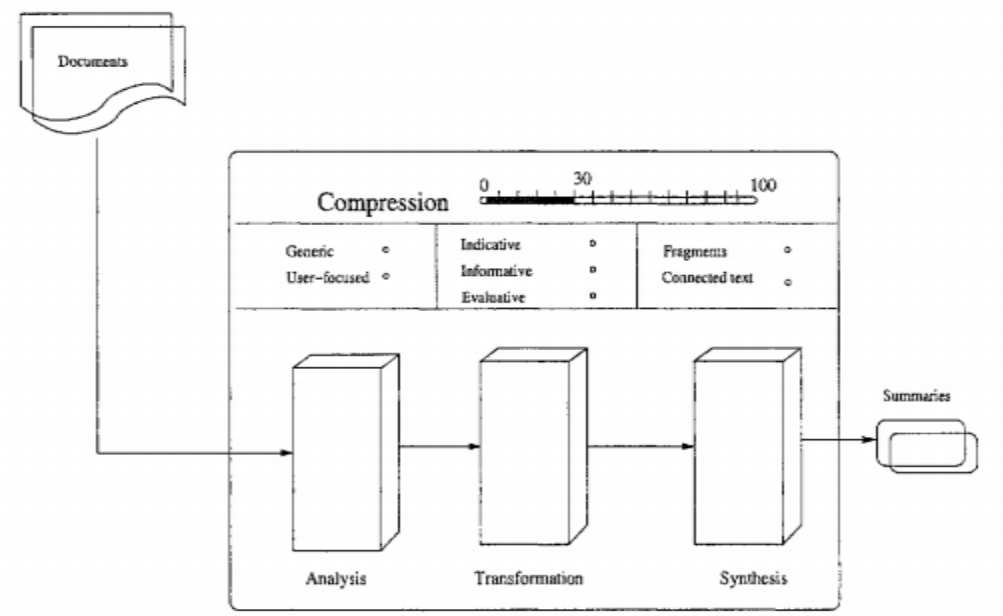

Figure 2.1 Diagrammatic representation of approach

These basic condensation operations can be applied during various phases of summarization on elements such as words, phrases, clauses, sentences, or discourse. Elements in these condensation operations can be analyzed at various linguistic levels: morphological, syntactic, and semantic and discourse/pragmatic. Based on the level of linguistic analysis of the source, summarization methods can be broadly classified into two approaches:

1. **Shallow approaches:** These methods tend to identify the salient portions of the text based on the surface level analysis of the document. These methods extract the sentences, considered salient, and then re-arrange them to form a coherent summary. Since these methods extract the complete sentence(s), they cannot achieve greater compression rates compared to the deeper approaches.

2. **Deeper approaches:** These methods perform deeper semantic analysis of the document content to identify the salient portions. They require highly domain-specific information to be able to perform deeper analysis. Lack of such widely available knowledge bases factors makes these methods hard to implement. One major advantage of these methods is the level of compression obtained.

CHAPTER 3

ANALYSIS

3. ANALYSIS

3.1 Lexical Chains

Lexical chains are sequence of semantically related words, spanning over the entire text. Consider the following example:

*Ammonia may have been found in **Mars'** atmosphere which some scientists say could indicate **life** on the **Red Planet**. The tentative detection of **ammonia** comes just a few months after **methane** was found in the **Martian** atmosphere. **Methane** is another **gas** with a possible **biological** origin.*

Hence Lexical chains can be constructed as:

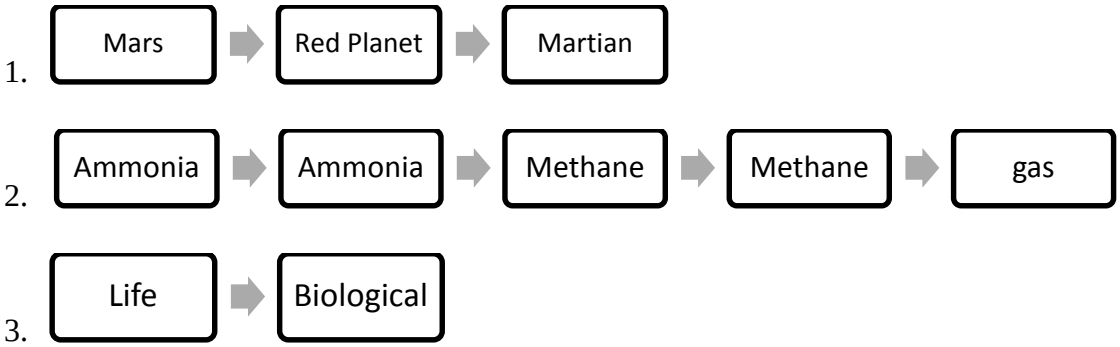

Figure 3.1 Lexical chains

Lexical chains can be computed by the surface level analysis of the text and would help us to identify the theme of the text (e.g.: "life on mars" for the above text). Lexical chains are used in various NLP applications; indexing for information retrieval, to correct malapropism, to divide the text into smaller segments, automatic hypertext construc-tion between two texts.

Lexical chains are also useful in identifying the sense of the word being used in the current context. For example, consider the word "bank" which can have two senses such as "a financial institution" or "river side". Given the lexical chain "{bank, slope, incline}", we can narrow down the sense of the word "bank" being used in this context to the "river side". This process of identifying the sense of the word in the given context is called as "word sense disambiguation" (WSD). WSD is important to identify the topic or theme of the document and is helpful in various tasks: summarization, query processing, text similarity, etc.

3.2 Why Lexical chain algorithm is used?

When deciding which algorithm should be used for this project, a vast amount of different algorithms were studied, ranging from surface-level approaches, entity-level to discourse-level approaches as mentioned earlier: Summaries can be built on a deep semantic analysis of the text. For example in McKeown and Radev investigate ways to produce a coherent summary of several texts describing the same event, when a full semantic representation of the source texts is available. This type of abstraction is the most expressive, yet very domain dependent and computational power is demanding. On the other hand, summaries can be built from a shallow linguistic analysis of the text. From all journals Barzilay and Elhadad studied, they concluded that all the techniques presented are easily computed and rely on formal clues found in the text (e.g. word frequencies, title words, location, cue phrases, sentence length, etc). As reported in (Paice 1990), location and cue phrases produced better results then the word frequency method, and can be accurately computed. However, Barzilay and Elhadad also pointed out that there are limitation on location and cue phrases method; that when the number of rhetorical markers changes critically, there are large difference in accuracy. Techniques relying on formal clues can be seen as a high gamble.

Method that rely more on the content do not suffer from this brittleness, and Lexical Chains is an example of such a method. The method Barzilay and Elhadad presented rely on word distribution and lexical links among them to approximate content in a more robust form. This produces a summary of an original text without requiring its full semantic interpretation, but instead relies on a model of topic progression in the text derived from lexical chains.

Lexical cohesion is defined as the cohesion that arises from the semantic relations between the words in the text. Lexical cohesion provides a good indicator for the discourse structure of the text, used by professional abstractors to skim through the document. Lexical cohesion does not occur just between two words but a sequence of related words spanning the entire text,

Lexical chains are used in a variety of NLP and IR applications such as summarization, detection of malapropism, indexing document for information retrieval, dividing the text into smaller segments based on the topic shift, automatic hypertext construction.

3.3 Chain computing

3.3.1 Generating Chain

Generally, a procedure for constructing lexical chains follows three steps:

1. Select a set of candidate words;
2. For each candidate word, find an appropriate chain relying on a relatedness criterion among members of the chains;
3. If it is found, insert the word in the chain and update it accordingly.

3.3.2 Building summary

Scoring Chains In order to use lexical chains as outlined above, one must first identify the strongest chains among all those that are produced by our algorithm.

3.3.3Extracting Significant Sentences

Once strong chains have been selected, the next step of the summarization algorithm is to extract full sentences from the original text based on chain distribution. We investigated three alternatives for this step:

3.4 Algorithm

Nouns, compound nouns and proper nouns as candidate words are considered to compute lexical chains. This is based on the intuition that nouns characterize the topic in the documents and that most of the documents describe a certain topic or a concept having various topics. Each candidate word is expanded to all of its senses.

Hash structure representation to identify all possible word representations are created. Each word sense is inserted into the hash entry having the index value equal to its synsetlD. For example, celebration and jubilation are inserted into the same hash entry in following figure.

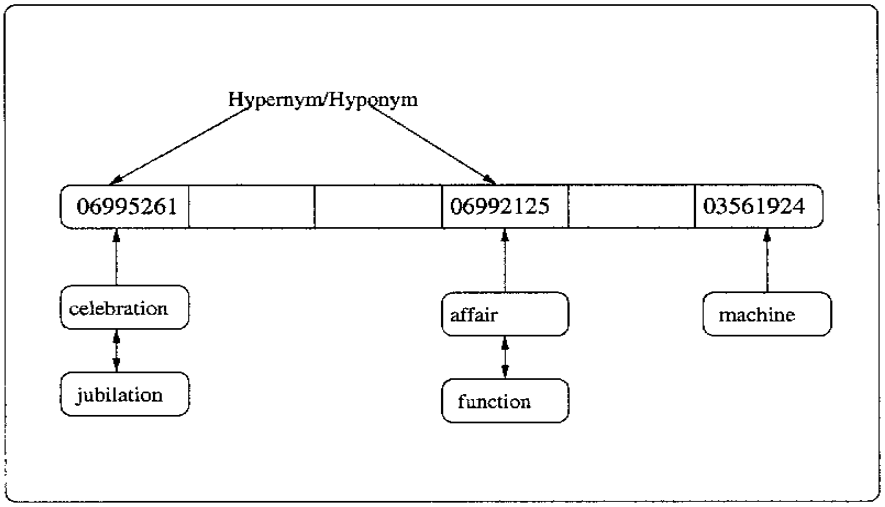

Figure 3.2 Hash structure indexed by synsetlD value

On insertion of the candidate sense into the hash structure, we check to see if there exists an entry into the index value, with which the current word sense has one of the following relations:
For each candidate sense inserted, we check to see if it is related (semantically) with any of the already present members in the structure. The relations considered are:

- Identical relation:
  e.g.:- Weather is great in Atlanta. Florida is having a really bad weather.
- Synonym relation (words belonging to the same synset in WordNet):
  e.g.:- Not all criminals are outlaws.
- Hypernym/Hyponym relation:
  e.g.:- Peter bought a computer. It was a Dell machine.
- Siblings (If the words have the same hypernym):
  e.g.:- Ganges flows into the Bay of Bengal. Amazon flows into the South Atlantic.
- Gloss (If the concept is present in the gloss of the word):
  eg:- gloss of word "dormitory " is {a college or university building containing living quarters for students}

Each relation is scored based on the distance (dist) between the two concepts in WordNet hierarchy = I / (dist + 1)

Relation	**Score**
Identical	1
Synonym	1
Hypernym/Hyponym	0.5
Sibling	0.33
Gloss	0.4

Table 3.1 Score of each relation (based on the length of path in WordNet)

For each candidate word sense, we identify the chains in which there exists a relation with each and every member of the chain. If found, we insert the word sense into the chain and update the score of the chain. Chain score is computed as the sum of scores of each relation in the chain which also includes the repetition count of each word.

$$\text{Score (chain)} = \sum_{i=1}^{n} \text{Score } (R_i))$$

Where R_i is the semantic measure between two members of the lexical chain.
Once the chains are computed, we sort the chains based on their score to determine the strength of the chains. We then filter out the chains which are not compatible with the higher ranked chains (i.e. having word from a higher ranked chain used in different sense). We retain the rest of the chains, which do not have words used in different sense to ones already assigned by higher ranked chains. This process of retaining only certain chains enables us to disambiguate the sense of the word being used in a particular context. This property can be used to evaluate the efficiency of lexical chaining algorithms based on their efficiency to correctly disambiguate the sense of a word.

CHAPTER 4

DESIGN

4. DESIGN

4.1 Architecture of Summarizer

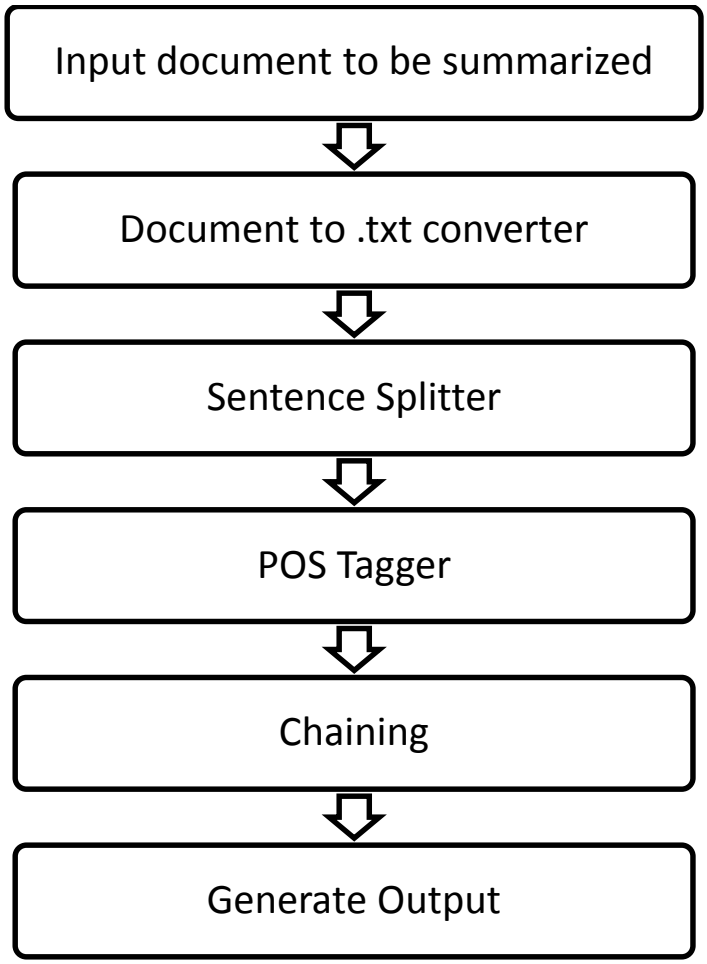

Figure 4.1 Architecture

4.1.1 Input document to be summarized

There are two ways in which input to summarizer application can be given, one is by using file browser and select the document to be summarized and other is writing text in input box provided.

4.1.2 Document to .txt converter

Input document can be in .docx, .doc, .pdf format which is converted to common .txt format. Converted file is in following format

India is a country in South Asia. It is the seventh-largest country by area, the second-most populous country with over 1.2 billion people, and the most populous democracy in the world. Bounded by the Indian Ocean on the south, the Arabian Sea on the south-west, and the Bay of Bengal on the south-east, it shares land borders with Pakistan to the west; China, Nepal, and Bhutan to the north-east; and Burma and Bangladesh to the east. In the Indian Ocean, India is in the vicinity of Sri Lanka and the Maldives; in addition, India's Andaman

4.1.3 Sentence splitter

Each sentence is written on new line and are tagged with sentence number. '▬' is used to represent new paragraph. Sentences are written in following format

◄1►India
▬
◄2►India is a country in South Asia.
◄3► It is the seventh-largest country by area, the second-most populous country with over
◄4►2 billion people, and the most populous democracy in the world.
◄5► Bounded by the Indian Ocean on the south, the Arabian Sea on the south-west, and the Bay of Bengal on the south-east, it shares land borders with Pakistan to the west; China,

4.1.4 POS Tagger

In this phase Stanford POS tagger is used to tag words with their Parts of Speech. Each sentence is given line by line to tagger. Word and its tag is separated by '/' Tagged output is as follows:

◄/NN 1/CD ►/CD India/NNP
▬/NN
◄/NN 2/CD ►/CD India/NNP is/VBZ a/DT country/NN in/IN South/NNP Asia/NNP ./.
◄/NN 3/CD ►/NN It/PRP is/VBZ the/DT seventh-largest/JJ country/NN by/IN area/NN ,/, the/DT second-most/JJ populous/JJ country/NN with/IN over/IN 1/CD ./.
◄/NN 4/CD ►/NN 2/CD billion/CD people/NNS ,/, and/CC the/DT most/RBS populous/JJ democracy/NN in/IN the/DT world/NN ./.

4.1.5 Chaining

In this phase nouns are taken for chaining. Each noun is searched in wordnet dictionary for synonyms, hyponyms, siblings, gloss. Similar words are chained and chain score is calculated. [63] Represents chain score and (2) represents score of word. Chains are generated as follows:

[63] India(2)--country(2)--Asia(2)--country(3)--area(3)--country(3)--people(4)--Indian(5)--Bengal(5)--land(5)--Pakistan(5)--China(5)--Nepal(5)--Bhutan(5)--Burma(5)--Bangladesh(5)--Indian(6)--India(6)--Maldives(6)--India(6)--Thailand(6)--Indonesia(6)

[28] South(2)--democracy(4)--Ocean(5)--south(5)--Sea(5)--south-west(5)--Bay(5)--south-east(5)--borders(5)--west(5)--north-east(5)--Ocean(6)--addition(6)--Islands(6)--border(6)

[1] world(4)--share(6)

4.1.6 Generate Output

Top scored chains are selected and sentences are generated by extracting sentences from input. Output is displayed in .txt format

India is a country in South Asia. It is the seventh-largest country by area, the second-most populous country with over 1.2 billion people, and the most populous democracy in the world. During the period 2000–500 BCE, in terms of culture, many regions of the subcontinent transitioned from the Chalcolithic to the Iron Age. The Vedas, the oldest scriptures of Hinduism, were composed during this period, and historians have analysed these to posit a Vedic culture in the Punjab region and the upper Gangetic Plain.

4.2 Flow Chart

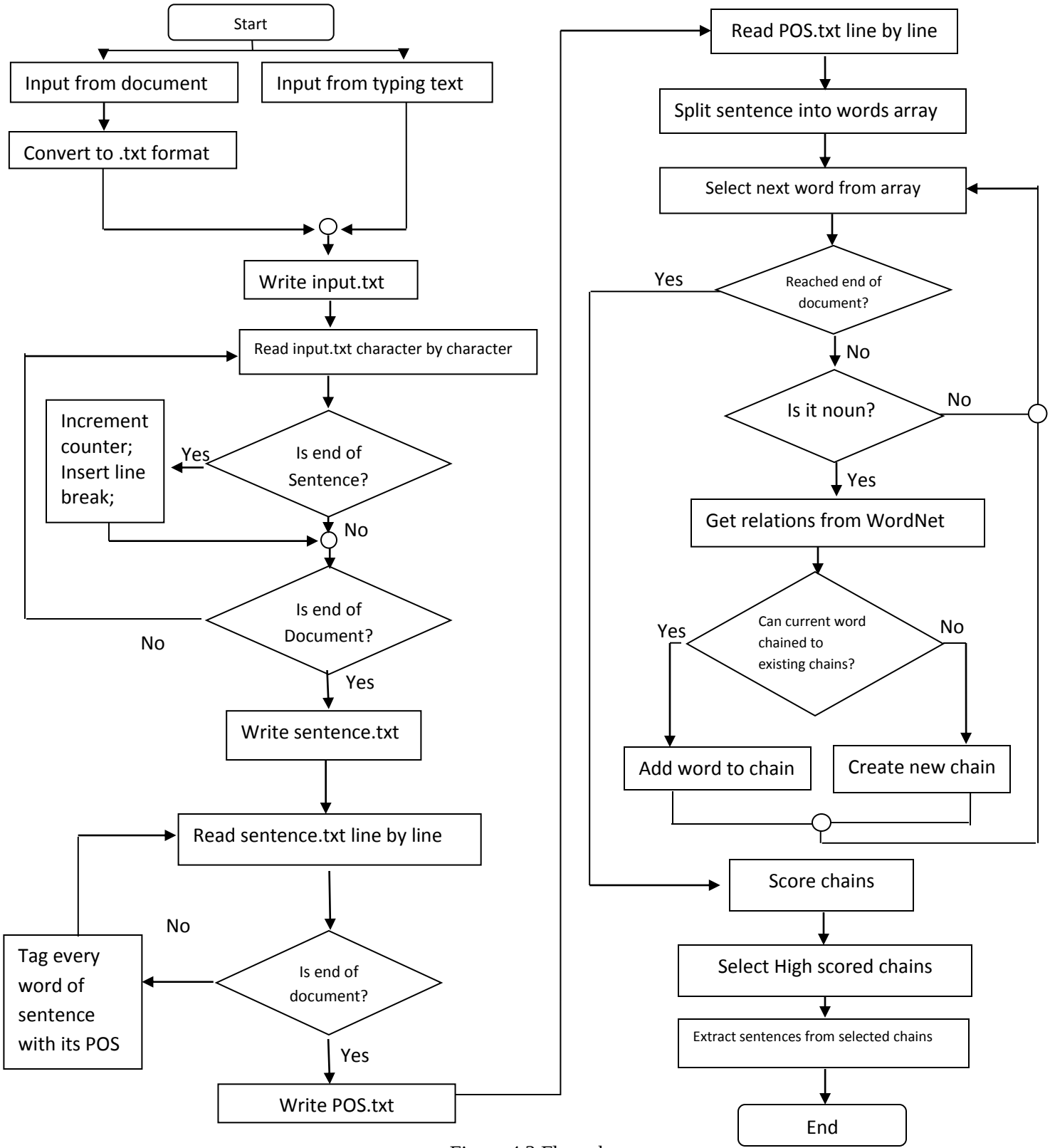

Figure 4.3 Flow chart

4.3 Gantt Chart

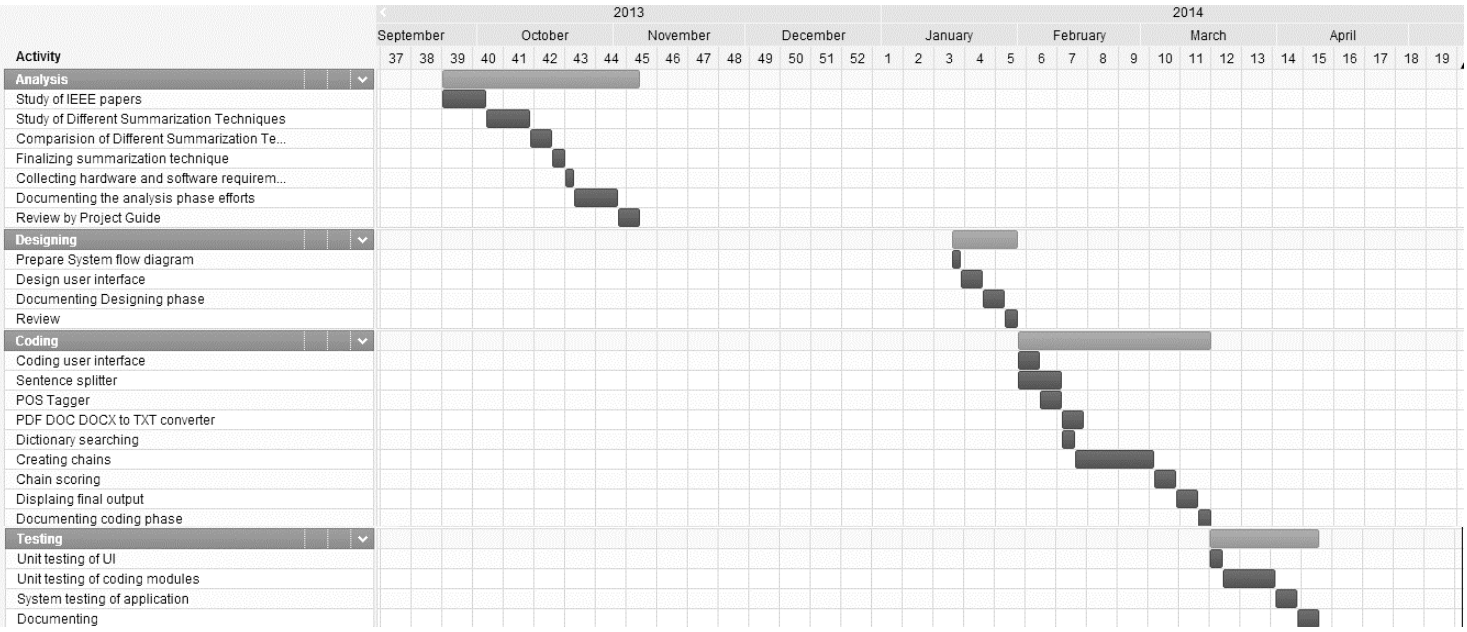

Figure 4.3 Gantt chart

CHAPTER 5

IMPLEMENTATION

5. Implementation

5.1 Text converter

```
private void jFileChooser1ActionPerformed(java.awt.event.ActionEvent
evt) {
        File f;
        String cnsl;
        f = jFileChooser1.getSelectedFile();
        percent = jSlider2.getValue();
        cnsl = "Opening document "+f.getName();
        cnsl+= "\n Summarizing to"+percent+"%";
        jTextPane1.setText(cnsl);
        System.out.println(f.getPath());
        String s="";
        FileInputStream fis = null;
        try{
        fis=new FileInputStream(f.getAbsolutePath());
        }catch(Exception e){e.printStackTrace();}
        String fileExtension= getExtension(f.getPath());
        if(fileExtension.equals("doc"))
        {
            try {
                s =
org.apache.poi.hwpf.converter.WordToTextConverter.getText(f);
            } catch (Exception ex) {

Logger.getLogger(Main.class.getName()).log(Level.SEVERE, null, ex);
            }
           writeFile(s);

        }
        else if(fileExtension.equals("docx"))
        {

WordToTextConverter.convertWordToText(f.getPath(),"input.txt");

        }
        else if(fileExtension.equals("pdf"))
        {
           PDFToTextConverter.convertPDFToText(f.getPath(),
"input.txt");
            try {
                PDFFilter.doFilter();
            } catch (Exception ex) {

Logger.getLogger(Main.class.getName()).log(Level.SEVERE, null, ex);
            }
        }
       System.out.println("Input converted to text");
```

```
        ju2 = new
DocumentBrowserSummaryThread(jProgressBar1,jTextPane1);
        Thread t2 = new Thread(ju2);
        t2.setPriority(Thread.NORM_PRIORITY);
        t2.start();
}

public String getExtension(String fileName){
         String extension = "";

         int i = fileName.lastIndexOf('.');
         int p = Math.max(fileName.lastIndexOf('/'),
fileName.lastIndexOf('\\'));

         if (i > p) {
             extension = fileName.substring(i+1);
         }
         return extension;
    }

    public static void writeFile(String input){
         try{
        OutputStream outputFile = new FileOutputStream("input.txt");
        byte write[] = input.getBytes();
         for(byte x: write){
             outputFile.write(x);
         }
        outputFile.close();
        }catch(Exception e){
            e.printStackTrace();
        }
    }
```

5.2 Sentence Splitter

```
public static void splitSentence() {
         System.out.println("Sentense spliting started...");
         BufferedReader br = null;
         String str = "◄1►";
         try {


             count = 1;
             char c;
             br = new BufferedReader(new FileReader("input.txt"));

             while (true) {
                 c = (char) br.read();
                 //System.out.print(c+"("+(int)c+")");
                 if(-1==(int)c||65535==(int)c)break;
```

```
                if (c == '.') {
                    count++;
                    str += c +
System.getProperty("line.separator")+'◄'+count+'►';

                    if (count % 100 == 0) {
                        System.out.println(count);
                    }
                }
                else if(c == (char)10){
                    count++;
                    str += c +"—"+
System.getProperty("line.separator")+'◄'+count+'►';

                    if (count % 100 == 0) {
                        System.out.println(count);
                    }
                }
                else{str += c;}
            }
            OutputStream outputFile = new
FileOutputStream("sentence.txt");
                byte write[] = str.getBytes();
                for (byte x : write) {
                    outputFile.write(x);
                }

                outputFile.close();
        } catch (IOException e) {
            e.printStackTrace();
        }

    }
```

5.3 Line Break Filter

```
public static void doFilter() throws FileNotFoundException,
IOException{
        System.out.println("pdf filter started...");
        BufferedReader  br;
        char c;
        String s;
        int total=0,sqTotal=0,noOfLines=1,col=0;
        int mean,variance,sd;

        String op="";


        //LINE BREAK FILTER
```

```
        br = new BufferedReader(new FileReader("input.txt"));
        while ((s = br.readLine()) != null) {
            noOfLines++;
            total += s.length();

        }
        mean = total/noOfLines;
        System.out.println("Mean: "+mean);

        br = new BufferedReader(new FileReader("input.txt"));
        while ((s = br.readLine()) != null) {
            sqTotal += Math.pow((mean-s.length()),2);
        }

        variance = sqTotal/noOfLines;
        sd = (int)Math.pow(variance, 0.5);
        int cutoff = mean - sd;
        int oldcutoff = cutoff;

        System.out.println("variance: "+variance+" sd: "+sd+" cutoff
"+cutoff);

        //NESTED FILTER
     for(int i=0; i<100; i++){
        System.out.println("Nested filter");
        br = new BufferedReader(new FileReader("input.txt"));
        while ((s = br.readLine()) != null) {
            if(s.length()>cutoff){
                noOfLines++;
                total += s.length();
            }
        }
        mean = total/noOfLines;
        System.out.println("Mean: "+mean);

        br = new BufferedReader(new FileReader("input.txt"));

        while ((s = br.readLine()) != null) {
            if(s.length()>cutoff){
                sqTotal += Math.pow((mean-s.length()),2);
            }
        }

        variance = sqTotal/noOfLines;
        sd = (int)Math.pow(variance, 0.5);
        cutoff = mean - sd;

        System.out.println("variance: "+variance+" sd: "+sd+" cutoff
"+cutoff);
        if(oldcutoff==cutoff){break;}
        oldcutoff = cutoff;
     }
```

```
        op="";
        br = new BufferedReader(new FileReader("input.txt"));
        while ((s = br.readLine()) != null) {
               if(s.length()<=cutoff){
                   op+=(s+System.getProperty("line.separator"));
                   //System.out.println("new para");
               }else{
                   op+=(s.substring(0, s.length()));
               }
        }
        Main.writeFile(op);
}
```

5.4 POS Tagger

```
class Line {

    String string;
    int line;

    Line(String s, int i) {
        string = s;
        line = i;
    }
}
public static synchronized Line getLine() {
        try {
            String ret = br.readLine();
            if (ret != null) {
                buffer.add(ret);
                countLines(ret);
                return new Line(ret, (buffer.size() - 1));
            }
            else{
                return null;
            }

        } catch (IOException ex) {

Logger.getLogger(POSMultiThreaded3.class.getName()).log(Level.SEVERE,
null, ex);
            return null;
        }
    }
    static void countLines(String line){
        if(line.startsWith("◄")){
            try{
            String[] s1 = line.split("◄");
            String[] s2 = s1[s1.length-1].split("►");
```

```
            sentences_done = Integer.parseInt(s2[0]);
            }catch(Exception e){e.printStackTrace();}
        }
    }

    public static synchronized void writeFile(){
        if(!completed){
            completed = true;
            System.out.println("Writing POS.txt");
            try{
            OutputStream outputFile = new FileOutputStream("pos.txt");
            String str="";
            for(int i=0;i<buffer.size();i++){
                str+=
buffer.get(i)+System.getProperty("line.separator");;
            }
            byte write[] = str.getBytes();
             for(byte x: write){
                 outputFile.write(x);
             }
             System.out.println("POS Tagging done");
            outputFile.close();
            POSMultiThreaded3.writingDone=true;
        }
            catch(Exception e){
                e.printStackTrace();
            }
        }
    }

    public static void doPOS() {
        try {
            br = new BufferedReader(new FileReader("sentence.txt"));
        } catch (FileNotFoundException ex) {

Logger.getLogger(POSMultiThreaded3.class.getName()).log(Level.SEVERE,
null, ex);
        }

        POSThread3 pos1 = new POSThread3();
        pos1.start();
         POSThread3 pos2 = new POSThread3();
        pos2.start();
         POSThread3 pos3 = new POSThread3();
        pos3.start();
         POSThread3 pos4 = new POSThread3();
        pos4.start();

        while(!writingDone){
            try {

//System.out.println(sentences_done+"/"+sentense.SimpleSpliter.count);
```

```
                if(summarizer.gui.InputTextSummaryThread.jpb!=null)

summarizer.gui.InputTextSummaryThread.jpb.setValue(15+(sentences_done*
50/sentense.SimpleSpliter.count));

if(summarizer.gui.DocumentBrowserSummaryThread.jpb!=null)

summarizer.gui.DocumentBrowserSummaryThread.jpb.setValue(15+(sentences
_done*50/sentense.SimpleSpliter.count));
                Thread.sleep(1000);
            } catch (InterruptedException ex) {

Logger.getLogger(POSMultiThreaded3.class.getName()).log(Level.SEVERE,
null, ex);
            }
        }
class POSThread3 extends Thread {
    private Thread t;
    MaxentTagger tagger;
    Line ln;
    POSThread3() {
         tagger =  new MaxentTagger("lib/models/english-bidirectional-
distsim.tagger"); //LOCATION OF MODEL
    }

    public void run() {
        try {
            String tagged,newTag;
           while((ln =  POSMultiThreaded3.getLine()) != null){
                 tagged = tagger.tagString(ln.string);
                  newTag= tagged.replace('_', '/');
                  //System.out.println(newTag);

                  POSMultiThreaded3.buffer.set(ln.line, newTag);
           }
           tagger = null;
           this.sleep(500);
           POSMultiThreaded3.writeFile();
        } catch (Exception e) {
            e.printStackTrace();
        }
    }

    public void start() {

        if (t == null) {
            t = new Thread(this);
            t.start();
        }
    }
}
```

5.5 Chaining

```
public static void doChaning() throws FileNotFoundException,
IOException{
        String sCurrentLine;
        String[] tokens, data;
        MyWord temp;
        //Chain tempChain = new Chain();
        //Chain tempChain2 = new Chain();
        Chain tempChain3 = new Chain();
        //chainTable.add(tempChain);
        boolean newWord3, newPara=false;
        int
count=0,line=0,ct3Lim=0,percent=summarizer.gui.Main.percent,totalLines
=sentense.SimpleSpliter.count,linesDone=0,total3;
        if(percent==0){percent=33;}
        BufferedReader br = new BufferedReader(new
FileReader("pos.txt"));
        System.out.println("Chaining Started...");
        try {
            wordnet.WordNetHandler.initialize();
        } catch (Exception ex) {
            System.err.print("Error loading dictionary");
        }
        while ((sCurrentLine = br.readLine()) != null) {
            tokens = sCurrentLine.split(" ");
            for (int i = 0 ; i < tokens.length ; ++i){
                data = tokens[i].split("/");
               // if(data.length==2){
                //System.out.println("["+data[0]+" "+data[1]+" ");
                //System.out.print((data.length==2)+"
"+(data[1].equalsIgnoreCase("NN"))+"] ");
                //}

if(data[0].contains("◄")){try{line=Integer.parseInt(tokens[i+1].split(
"/")[0]);}catch(Exception e){}}
                else if(data[0].contains("►")){}
                else if(data[0].contains("—")){newPara=true;}
                else if(data.length>1 &&
(data[1].equalsIgnoreCase("NN")||data[1].equalsIgnoreCase("NNS")||data
[1].equalsIgnoreCase("NNP")||data[1].equalsIgnoreCase("NNPS"))){
                     count++;
                       // if(count%25==0)System.out.println(count);
                    temp = new MyWord(data[0],line);  //TEMPORARILY
COMMENTED



                    //ALGORITHM 3 - CHAINTABLE 3
                                          //TEMPORARILY COMMENTED
                    newWord3 = true;
```

```
                    if(newPara){

                        ct3Lim=chainTable3.size();
                        //System.out.println("new para at
line="+line+", limit="+ct3Lim);
                        newPara=false;
                    }
                    for(int j=ct3Lim;j<chainTable3.size();j++){
                        if(chainTable3.get(j).add(temp)){
                           // System.out.println(temp.word+" chained @
"+j+", limit="+ct3Lim+", size="+chainTable3.size());
                           // chainTable3.get(j).print();
                           newWord3=false; break;
                        }
                    }
                    if(newWord3){
                        tempChain3 = new Chain();
                        tempChain3.words.add(temp);
                        chainTable3.add(tempChain3);
                    }
                 //

//System.out.println(temp.word+"\n"+temp.Synonym.toString()+temp.PartO
f.toString()+temp.Hyponym.toString()+temp.Hypernym.toString()+temp.Glo
ss.toString()+"\n----------------\n");
                }
            }

            //if(line%1==0){
                if(summarizer.gui.InputTextSummaryThread.jpb!=null)

summarizer.gui.InputTextSummaryThread.jpb.setValue(65+(line*25/totalLi
nes));

if(summarizer.gui.DocumentBrowserSummaryThread.jpb!=null)

summarizer.gui.DocumentBrowserSummaryThread.jpb.setValue(65+(line*25/t
otalLines));


       System.out.println("----------Printing chaintable3-----------
--");
        for(int i=0;i<chainTable3.size();i++){
           chainTable3.get(i).print();
           System.out.println();
       }
```

5.6 Searching word in WordNet dictionary

```
public static void initialize()throws JWNLException,
FileNotFoundException{
         JWNL.initialize(new FileInputStream("file_properties.xml"));
         dictionary = Dictionary.getInstance();
    }
    public static void getRelation(MyWord wrd)throws JWNLException,
FileNotFoundException{

        final IndexWord indexWord =
dictionary.lookupIndexWord(POS.NOUN, wrd.word);

        try{
            final Synset[] senses = indexWord.getSenses();

         Synset[] synSets = indexWord.getSenses();

         for (Synset synset : synSets){
            Word[] words = synset.getWords();
            for (Word word : words){
               wrd.Synonym.add(word.getLemma());
            }
         }

         for (Synset synset : synSets)
         {
               PointerTarget[] targets =
synset.getTargets(PointerType.HYPERNYM);
               for (PointerTarget target : targets){
                  Word[] words = ((Synset) target).getWords();
                  for (Word word : words){
                     String tmp[] = word.getLemma().split("_");
                     for(String x:tmp){wrd.Hypernym.add(x);}

                  }
               }
         }



         for (Synset synset : synSets){
               PointerTarget[] targets =
synset.getTargets(PointerType.HYPONYM);
               for (PointerTarget target : targets){
                  Word[] words = ((Synset) target).getWords();
                  for (Word word : words){
                      String tmp[] = word.getLemma().split("_");
```

```
                        for(String x:tmp){wrd.Hyponym.add(x);}
                    }
                }
            }
                    //
                     for (int i=0; i<senses.length; i++) {
                    Synset sense = senses[i];
                    //System.out.println((i+1) + ". " + sense.getGloss());
                    wrd.Gloss.add(sense.getGloss());
                    Pointer[] holo =
sense.getPointers(PointerType.PART_HOLONYM);
                    for (int j=0; j<holo.length; j++) {
                      Synset synset = (Synset) (holo[j].getTarget());
                      Word synsetWord = synset.getWord(0);
                      wrd.PartOf.add(synsetWord.getLemma());
                      //System.out.print("  -part-of-> " +
synsetWord.getLemma());
                      //System.out.println(" = " + synset.getGloss());
                    }
                }
            }catch(NullPointerException e){ }

    }


class MyWord {
    public String word;
    public int lineNo;
    public MyWord(String w, int l){
        try {
            word = w;
            lineNo = l;
            WordNetHandler.getRelation(this);
        } catch (JWNLException | FileNotFoundException ex) {
            Logger.getLogger(MyWord.class.getName()).log(Level.SEVERE,
null, ex);
        }
    }
    public ArrayList<String>Synonym = new ArrayList<>();
    public ArrayList<String>Hypernym = new ArrayList<>();
    public ArrayList<String>Hyponym = new ArrayList<>();
    public ArrayList<String>PartOf = new ArrayList<>();
    public ArrayList<String>Gloss = new ArrayList<>();


    public boolean isSame(MyWord w){
        if(w.word.equalsIgnoreCase(word)) return true;
        else return false;
    }

    public boolean isSynonym(MyWord w){
        if(Synonym.contains(w.word)) return true;
```

```
    else if (w.Synonym.contains(this.word)) return true;
    else return false;
}


public boolean isHypernym(MyWord w){
    return (Hypernym.contains(w.word)||w.Hypernym.contains(word));
}


  public boolean isHyponym(MyWord w){
    return (Hyponym.contains(w.word)||w.Hyponym.contains(word));
}


public boolean isPartOf(MyWord w){
    return (PartOf.contains(w.word)||w.PartOf.contains(word));
}

 public boolean isGloss(MyWord w){
    for(int i=0; i<Gloss.size(); i++){
         String s[] = Gloss.get(i).split(" ");
         for(String x:s){
             if(x.equals(w.word)) return true;
         }
    }
    for(int i=0; i<w.Gloss.size(); i++){
         String s[] = w.Gloss.get(i).split(" ");
         for(String x:s){
             if(x.equals(word)) return true;
         }
    }
    return false;
}

public boolean isSybling(MyWord w){
    for(String x:Synonym){
        for(String y:w.Synonym){
            if(x.equalsIgnoreCase(y)) return true;
        }
        for(String y:w.PartOf){
            if(x.equalsIgnoreCase(y)) return true;
        }
        for(String y:w.Hypernym){
            if(x.equalsIgnoreCase(y)) return true;
        }
        for(String y:w.Hyponym){
            if(x.equalsIgnoreCase(y)) return true;
        }
    }

    for(String x:PartOf){
```

```
            for(String y:w.Synonym){
                if(x.equalsIgnoreCase(y)) return true;
            }
            for(String y:w.PartOf){
                if(x.equalsIgnoreCase(y)) return true;
            }
            for(String y:w.Hypernym){
                if(x.equalsIgnoreCase(y)) return true;
            }
            for(String y:w.Hyponym){
                if(x.equalsIgnoreCase(y)) return true;
            }
        }

        for(String x:Hyponym){
            for(String y:w.Synonym){
                if(x.equalsIgnoreCase(y)) return true;
            }
            for(String y:w.PartOf){
                if(x.equalsIgnoreCase(y)) return true;
            }
            for(String y:w.Hypernym){
                if(x.equalsIgnoreCase(y)) return true;
            }
            for(String y:w.Hyponym){
                if(x.equalsIgnoreCase(y)) return true;
            }
        }

        for(String x:Hypernym){
            for(String y:w.Synonym){
                if(x.equalsIgnoreCase(y)) return true;
            }
            for(String y:w.PartOf){
                if(x.equalsIgnoreCase(y)) return true;
            }
            for(String y:w.Hypernym){
                if(x.equalsIgnoreCase(y)) return true;
            }
            for(String y:w.Hyponym){
                if(x.equalsIgnoreCase(y)) return true;
            }
        }
        return false;
    }

    public boolean isSynonymSybling(MyWord w){
        for(String x:w.Synonym){
            for(String y:Synonym){
                if(x.equalsIgnoreCase(y)){ return true;}
            }
        }
```

```
        return false;
    }
}
```

5.7 Generating Output

```
displayOutput(){
     //SORTING
        int cutOff = count*percent/100;
        int actual = 0,actual2 = 0,actual3 = 0,i,j;

        ArrayList<ChainIndex> sorted3 = new ArrayList<>();
        for(int k=0;k<chainTable3.size();k++){
            if(sorted3.isEmpty()){sorted3.add(new
ChainIndex(chainTable3.get(k).score,k));}
            else{
                for(int l=0;l<sorted3.size();l++){
                    if(sorted3.get(l).score<chainTable3.get(k).score){
                        sorted3.add(l,new
ChainIndex(chainTable3.get(k).score,k));break;
                    }
                }
            }
        }

        System.out.println("sorting done");
      if(summarizer.gui.InputTextSummaryThread.jpb!=null)
        summarizer.gui.InputTextSummaryThread.jpb.setValue(95);
      if(summarizer.gui.DocumentBrowserSummaryThread.jpb!=null)
        summarizer.gui.DocumentBrowserSummaryThread.jpb.setValue(95);
       //CHAIN SELECTION
       String op="";

           ArrayList<Integer> selected3 = new ArrayList<>();

           for(ChainIndex x:sorted3){
               for(MyWord mw : chainTable3.get(x.index).words){
                    if(!selected3.contains(mw.lineNo)){
                        selected3.add(mw.lineNo);

                    }
                    actual3++;
                }
               if(actual3>cutOff)break;
           }
           Collections.sort(selected3);

           System.out.println("Printing selected3\n"+selected3);
           System.out.println("Summarized to "+percent+"%");
```

```
            String fin="",fin2="",fin3="";
            //SENTENCE LINKING
         String str = "",str2="",str3="";
         try {
             int lnno,lnno2,lnno3;
              br = new BufferedReader(new FileReader("sentence.txt"));
             while((str = br.readLine())!=null){
                 str2=str3=str;
                try{

                     str3 = str3.substring(1, str3.length());
                 lnno3 = Integer.parseInt(str3.split("►")[0]);
                 if(selected3.contains(lnno3)){
                     fin3+=str3.split("►")[1];
                 }


                 }catch(NumberFormatException e){}
             }

        OutputStream outputFile = new FileOutputStream("summary3.txt");
        byte write3[] = fin3.getBytes();
         for(byte x: write3){
             outputFile.write(x);
         }
        outputFile.close();
        summarizer.gui.Main.summary = fin3;
         }catch(Exception e){
             System.out.println(e);
         }

    }

}
```

CHAPTER 6

RESULTS

6. **Results**

6.1 Screenshots

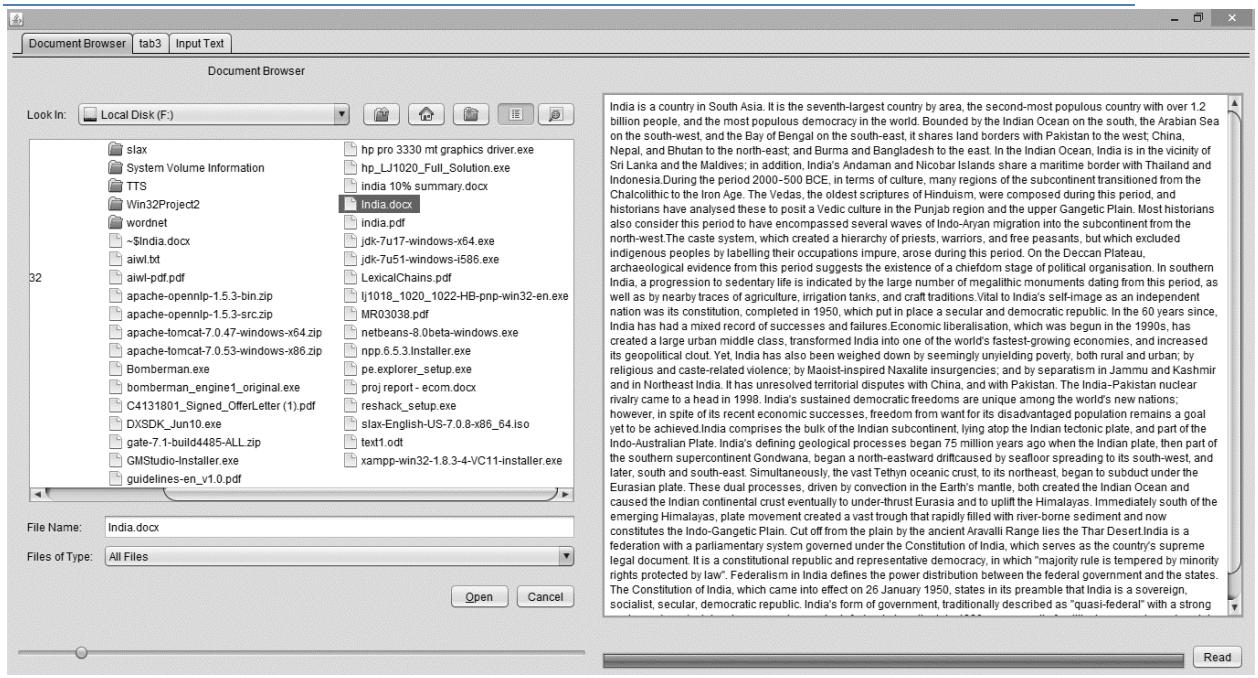

Image 7.1 Document summarizer window

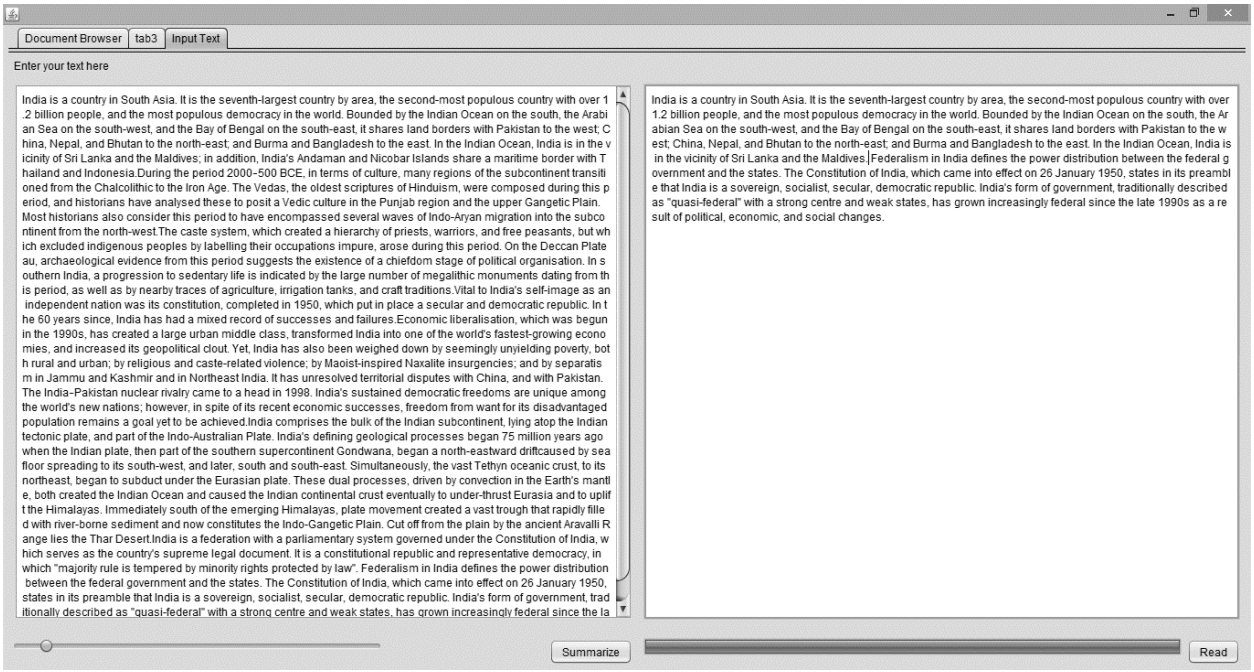

Image 7.2 Input text summarizer window

6.2 Comparison with existing systems

6.2.1 Input given (Article of India from Wikipedia):

India

India is a country in South Asia. It is the seventh-largest country by area, the second-most populous country with over 1.2 billion people, and the most populous democracy in the world. Bounded by the Indian Ocean on the south, the Arabian Sea on the south-west, and the Bay of Bengal on the south-east, it shares land borders with Pakistan to the west; China, Nepal, and Bhutan to the north-east; and Burma and Bangladesh to the east. In the Indian Ocean, India is in the vicinity of Sri Lanka and the Maldives; in addition, India's Andaman and Nicobar Islands share a maritime border with Thailand and Indonesia.

Home to the ancient Indus Valley Civilisation and a region of historic trade routes and vast empires, the Indian subcontinent was identified with its commercial and cultural wealth for much of its long history. Four world religions—Hinduism, Buddhism, Jainism, and Sikhism—originated here, whereas Judaism, Zoroastrianism, Christianity, and Islam arrived in the 1st millennium CE and also helped shape the region's diverse culture. Gradually annexed by and brought under the administration of the British East India Company from the early 18th century and administered directly by the United Kingdom from the mid-19th century, India became an independent nation in 1947 after a struggle for independence that was marked by non-violent resistance led by Mahatma Gandhi.

The Indian economy is the world's eleventh-largest by nominal GDP and third-largest by purchasing power parity (PPP). Following market-based economic reforms in 1991, India became one of the fastest-growing major economies; it is considered a newly industrialised country. However, it continues to face the challenges of poverty, corruption, malnutrition, inadequate public healthcare, and terrorism. A nuclear weapons state and a regional power, it has the third-largest standing army in the world and ranks eighth in military expenditure among nations. India is a federal constitutional republic governed under a parliamentary system consisting of 28 states and 7 union territories. India is a pluralistic, multilingual, and a multi-ethnic society. It is also home to a diversity of wildlife in a variety of protected habitats.

History

Ancient India

Anatomically modern humans are thought to have arrived in South Asia 73-55,000 years back,though the earliest authenticated human remains date to only about 30,000 years ago.

Nearly contemporaneous Mesolithic rock art sites have been found in many parts of the Indian subcontinent, including at the Bhimbetka rock shelters in Madhya Pradesh. Around 7000 BCE, the first known Neolithic settlements appeared on the subcontinent in Mehrgarh and other sites in western Pakistan. These gradually developed into the Indus Valley Civilisation, the first urban culture in South Asia;[26] it flourished during 2500–1900 BCE in Pakistan and western India. Centred around cities such as Mohenjo-daro, Harappa, Dholavira, and Kalibangan, and relying on varied forms of subsistence, the civilisation engaged robustly in crafts production and wide-ranging trade.

During the period 2000–500 BCE, in terms of culture, many regions of the subcontinent transitioned from the Chalcolithic to the Iron Age. The Vedas, the oldest scriptures of Hinduism, were composed during this period, and historians have analysed these to posit a Vedic culture in the Punjab region and the upper Gangetic Plain. Most historians also consider this period to have encompassed several waves of Indo-Aryan migration into the subcontinent from the north-west.The caste system, which created a hierarchy of priests, warriors, and free peasants, but which excluded indigenous peoples by labelling their occupations impure, arose during this period. On the Deccan Plateau, archaeological evidence from this period suggests the existence of a chiefdom stage of political organisation. In southern India, a progression to sedentary life is indicated by the large number of megalithic monuments dating from this period, as well as by nearby traces of agriculture, irrigation tanks, and craft traditions.

Early modern India

writing the will and testament of the Mughal king court in Persian, 1590–1595

In the early 16th century, northern India, being then under mainly Muslim rulers, fell again to the superior mobility and firepower of a new generation of Central Asian warriors. The resulting Mughal Empire did not stamp out the local societies it came to rule, but rather balanced and pacified them through new administrative practices and diverse and inclusive ruling elites, leading to more systematic, centralised, and uniform rule. Eschewing tribal bonds and Islamic identity, especially under Akbar, the Mughals united their far-flung realms through loyalty, expressed through a Persianised culture, to an emperor who had near-divine status. The Mughal state's economic policies, deriving most revenues from agriculture and mandating that taxes be paid in the well-regulated silver currency, caused peasants and artisans to enter larger markets. The relative peace maintained by the empire during much of the 17th century was a factor in India's economic expansion, resulting in greater patronage of painting, literary forms, textiles, and architecture.Newly coherent social groups in northern and western India, such as the Marathas, the Rajputs, and the Sikhs, gained military and governing ambitions during Mughal rule, which, through collaboration or adversity, gave them both recognition and military experience.Expanding commerce during Mughal rule gave rise to new Indian commercial and political elites along the coasts of southern and eastern India. As the empire disintegrated, many among these elites were able to seek and control their own affairs.

By the early 18th century, with the lines between commercial and political dominance being increasingly blurred, a number of European trading companies, including the English East India Company, had established coastal outposts. The East India Company's control of the seas, greater resources, and more advanced military training and technology led it to increasingly flex its military muscle and caused it to become attractive to a portion of the Indian elite; both these factors were crucial in allowing the Company to gain control over the Bengal region by 1765 and sideline the other European companies.Its further access to the riches of Bengal and the subsequent increased strength and size of its army enabled it to annex or subdue most of India by the 1820s.India was then no longer exporting manufactured goods as it long had, but was instead supplying the British empire with raw materials, and many historians consider this to be the onset of India's colonial period. By this time, with its economic power severely curtailed by the British parliament and itself effectively made an arm of British administration, the Company began to more consciously enter non-economic arenas such as education, social reform, and culture.

Modern India

The British Indian Empire, from the 1909 edition of The Imperial Gazetteer of India. Areas directly governed by the British are shaded pink; the princely states under British suzerainty are in yellow.

Historians consider India's modern age to have begun sometime between 1848 and 1885. The appointment in 1848 of Lord Dalhousie as Governor General of the East India Company set the stage for changes essential to a modern state. These included the consolidation and demarcation of sovereignty, the surveillance of the population, and the education of citizens. Technological changes—among them, railways, canals, and the telegraph—were introduced not long after their introduction in Europe. However, disaffection with the Company also grew during this time, and set off the Indian Rebellion of 1857. Fed by diverse resentments and perceptions, including invasive British-style social reforms, harsh land taxes, and summary treatment of some rich landowners and princes, the rebellion rocked many regions of northern and central India and shook the foundations of Company rule.Although the rebellion was suppressed by 1858, it led to the dissolution of the East India Company and to the direct administration of India by the British government. Proclaiming a unitary state and a gradual but limited British-style parliamentary system, the new rulers also protected princes and landed gentry as a feudal safeguard against future unrest.In the decades following, public life gradually emerged all over India, leading eventually to the founding of the Indian National Congress in 1885.

Two smiling men in robes sitting on the ground with bodies facing the viewer and with heads turned toward each other. The younger wears a white Nehru cap; the elder is bald and wears glasses. A half-dozen other people are in the background.

Jawaharlal Nehru (left) became India's first prime minister in 1947. Mahatma Gandhi (right) led the independence movement.

The rush of technology and the commercialisation of agriculture in the second half of the 19th century was marked by economic setbacks—many small farmers became dependent on the whims of far-away markets. There was an increase in the number of large-scale famines, and, despite the risks of infrastructure development borne by Indian taxpayers, little industrial employment was generated for Indians. There were also salutary effects: commercial cropping, especially in the newly canalled Punjab, led to increased food production for internal consumption. The railway network provided critical famine relief, notably reduced the cost of moving goods, and helped nascent Indian-owned industry.After World War I, in which some one million Indians served, a new period began. It was marked by British reforms but also repressive legislation, by more strident Indian calls for self-rule, and by the beginnings of a non-violent movement of non-cooperation, of which Mohandas Karamchand Gandhi would become the leader and enduring symbol. During the 1930s, slow legislative reform was enacted by the British; the Indian National Congress won victories in the resulting elections. The next decade was beset with crises: Indian participation in World War II, the Congress's final push for non-cooperation, and an upsurge of Muslim nationalism. All were capped by the advent of independence in 1947, but tempered by the partition of India into two states: India and Pakistan.

Vital to India's self-image as an independent nation was its constitution, completed in 1950, which put in place a secular and democratic republic. In the 60 years since, India has had a mixed record of successes and failures. It has remained a democracy with civil liberties, an activist Supreme Court, and a largely independent press.Economic liberalisation, which was begun in the 1990s, has created a large urban middle class, transformed India into one of the world's fastest-growing economies, and increased its geopolitical clout. Indian movies, music, and spiritual teachings play an increasing role in global culture. Yet, India has also been weighed down by seemingly unyielding poverty, both rural and urban; by religious and caste-related violence; by Maoist-inspired Naxalite insurgencies; and by separatism in Jammu and Kashmir and in Northeast India. It has unresolved territorial disputes with China, and with Pakistan. The India–Pakistan nuclear rivalry came to a head in 1998. India's sustained democratic freedoms are unique among the world's new nations; however, in spite of its recent economic successes, freedom from want for its disadvantaged population remains a goal yet to be achieved.

Geography

India comprises the bulk of the Indian subcontinent, lying atop the Indian tectonic plate, and part of the Indo-Australian Plate. India's defining geological processes began 75 million years ago when the Indian plate, then part of the southern supercontinent Gondwana, began a north-eastward driftcaused by seafloor spreading to its south-west, and later, south and south-east. Simultaneously, the vast Tethyn oceanic crust, to its northeast, began to subduct under the Eurasian plate. These dual processes, driven by convection in the Earth's mantle, both created the

Indian Ocean and caused the Indian continental crust eventually to under-thrust Eurasia and to uplift the Himalayas. Immediately south of the emerging Himalayas, plate movement created a vast trough that rapidly filled with river-borne sediment and now constitutes the Indo-Gangetic Plain. Cut off from the plain by the ancient Aravalli Range lies the Thar Desert.

Major Himalayan-origin rivers that substantially flow through India include the Ganges and the Brahmaputra, both of which drain into the Bay of Bengal.Important tributaries of the Ganges include the Yamuna and the Kosi; the latter's extremely low gradient often leads to severe floods and course changes.Major peninsular rivers, whose steeper gradients prevent their waters from flooding, include the Godavari, the Mahanadi, the Kaveri, and the Krishna, which also drain into the Bay of Bengal; and the Narmada and the Tapti, which drain into the Arabian Sea. Coastal features include the marshy Rann of Kutch of western India and the alluvial Sundarbans delta of eastern India; the latter is shared with Bangladesh. India has two archipelagos: the Lakshadweep, coral atolls off India's south-western coast; and the Andaman and Nicobar Islands, a volcanic chain in the Andaman Sea.

The Indian climate is strongly influenced by the Himalayas and the Thar Desert, both of which drive the economically and culturally pivotal summer and wintermonsoons. The Himalayas prevent cold Central Asian katabatic winds from blowing in, keeping the bulk of the Indian subcontinent warmer than most locations at similar latitudes. The Thar Desert plays a crucial role in attracting the moisture-laden south-west summer monsoon winds that, between June and October, provide the majority of India's rainfall. Four major climatic groupings predominate in India: tropical wet, tropical dry, subtropical humid, and montane.

Environment

In India, major environmental issues include forest and agricultural degradation of land; depletion of resources such as water, minerals, forest, sand, and rocks; environmental degradation; public health issues; loss of biodiversity; loss of resilience in ecosystems; and livelihood security for the poor. According to data collection and environment assessment studies of World Bank experts, between 1995 and 2010, the progress India has made in addressing its environmental issues and improving its environmental quality has been among the fastest in the world.

Government

India is a federation with a parliamentary system governed under the Constitution of India, which serves as the country's supreme legal document. It is a constitutional republic and representative democracy, in which "majority rule is tempered by minority rights protected by law". Federalism in India defines the power distribution between the federal government and the states. The government abides by constitutional checks and balances. The Constitution of India, which came into effect on 26 January 1950, states in its preamble that India is a sovereign, socialist, secular, democratic republic. India's form of government, traditionally described as

"quasi-federal" with a strong centre and weak states, has grown increasingly federal since the late 1990s as a result of political, economic, and social changes.

Subdivisions

India is a federation composed of 28 states and 7 union territories.All states, as well as the union territories of Puducherry and the National Capital Territory of Delhi, have elected legislatures and governments, both patterned on the Westminster model. The remaining five union territories are directly ruled by the centre through appointed administrators. In 1956, under the States Reorganisation Act, states were reorganised on a linguistic basis. Since then, their structure has remained largely unchanged. Each state or union territory is further divided into administrative districts. The districts in turn are further divided into tehsils and ultimately into villages.

6.2.2 Output obtained (Comparison with existing systems)

6.2.2.1 Our system

India is a country in South Asia. It is the seventh-largest country by area, the second-most populous country with over 1.2 billion people, and the most populous democracy in the world. Bounded by the Indian Ocean on the south, the Arabian Sea on the south-west, and the Bay of Bengal on the south-east, it shares land borders with Pakistan to the west; China, Nepal, and Bhutan to the north-east; and Burma and Bangladesh to the east. In the Indian Ocean, India is in the vicinity of Sri Lanka and the Maldives; in addition, India's Andaman and Nicobar Islands share a maritime border with Thailand and Indonesia.During the period 2000–500 BCE, in terms of culture, many regions of the subcontinent transitioned from the Chalcolithic to the Iron Age. The Vedas, the oldest scriptures of Hinduism, were composed during this period, and historians have analysed these to posit a Vedic culture in the Punjab region and the upper Gangetic Plain. Most historians also consider this period to have encompassed several waves of Indo-Aryan migration into the subcontinent from the north-west.The caste system, which created a hierarchy of priests, warriors, and free peasants, but which excluded indigenous peoples by labelling their occupations impure, arose during this period. On the Deccan Plateau, archaeological evidence from this period suggests the existence of a chiefdom stage of political organisation. In southern India, a progression to sedentary life is indicated by the large number of megalithic monuments dating from this period, as well as by nearby traces of agriculture, irrigation tanks, and craft traditions.Vital to India's self-image as an independent nation was its constitution, completed in 1950, which put in place a secular and democratic republic. In the 60 years since, India has had a mixed record of successes and failures.Economic liberalisation, which was begun in the 1990s, has created a large urban middle class, transformed India into one of the world's fastest-growing

economies, and increased its geopolitical clout. Yet, India has also been weighed down by seemingly unyielding poverty, both rural and urban; by religious and caste-related violence; by Maoist-inspired Naxalite insurgencies; and by separatism in Jammu and Kashmir and in Northeast India. It has unresolved territorial disputes with China, and with Pakistan. The India–Pakistan nuclear rivalry came to a head in 1998. India's sustained democratic freedoms are unique among the world's new nations; however, in spite of its recent economic successes, freedom from want for its disadvantaged population remains a goal yet to be achieved.India comprises the bulk of the Indian subcontinent, lying atop the Indian tectonic plate, and part of the Indo-Australian Plate. India's defining geological processes began 75 million years ago when the Indian plate, then part of the southern supercontinent Gondwana, began a north-eastward driftcaused by seafloor spreading to its south-west, and later, south and south-east. Simultaneously, the vast Tethyn oceanic crust, to its northeast, began to subduct under the Eurasian plate. These dual processes, driven by convection in the Earth's mantle, both created the Indian Ocean and caused the Indian continental crust eventually to under-thrust Eurasia and to uplift the Himalayas. Immediately south of the emerging Himalayas, plate movement created a vast trough that rapidly filled with river-borne sediment and now constitutes the Indo-Gangetic Plain. Cut off from the plain by the ancient Aravalli Range lies the Thar Desert.India is a federation with a parliamentary system governed under the Constitution of India, which serves as the country's supreme legal document. It is a constitutional republic and representative democracy, in which "majority rule is tempered by minority rights protected by law". Federalism in India defines the power distribution between the federal government and the states. The Constitution of India, which came into effect on 26 January 1950, states in its preamble that India is a sovereign, socialist, secular, democratic republic. India's form of government, traditionally described as "quasi-federal" with a strong centre and weak states, has grown increasingly federal since the late 1990s as a result of political, economic, and social changes.

6.2.2.1 Text Compactor

Gradually annexed by and brought under the administration of the British East India Company from the early 18th century and administered directly by the United Kingdom from the mid-19th century, India became an independent nation in 1947 after a struggle for independence that was marked by non-violent resistance led by Mahatma Gandhi.

The Indian economy is the world's eleventh-largest by nominal GDP and third-largest by purchasing power parity (PPP). In southern India, a progression to sedentary life is indicated by the large number of megalithic monuments dating from this period, as well as by nearby traces of agriculture, irrigation tanks, and craft traditions.

Early modern India

writing the will and testament of the Mughal king court in Persian, 1590–1595

In the early 16th century, northern India, being then under mainly Muslim rulers, fell again to the superior mobility and firepower of a new generation of Central Asian warriors. The relative peace maintained by the empire during much of the 17th century was a factor in India's economic expansion, resulting in greater patronage of painting, literary forms, textiles, and architecture. Newly coherent social groups in northern and western India, such as the Marathas, the Rajputs, and the Sikhs, gained military and governing ambitions during Mughal rule, which, through collaboration or adversity, gave them both recognition and military experience. Expanding commerce during Mughal rule gave rise to new Indian commercial and political elites along the coasts of southern and eastern India. both these factors were crucial in allowing the Company to gain control over the Bengal region by 1765 and sideline the other European companies. Its further access to the riches of Bengal and the subsequent increased strength and size of its army enabled it to annex or subdue most of India by the 1820s.India was then no longer exporting manufactured goods as it long had, but was instead supplying the British empire with raw materials, and many historians consider this to be the onset of India's colonial period. Fed by diverse resentments and perceptions, including invasive British-style social reforms, harsh land taxes, and summary treatment of some rich landowners and princes, the rebellion rocked many regions of northern and central India and shook the foundations of Company rule. Although the rebellion was suppressed by 1858, it led to the dissolution of the East India Company and to the direct administration of India by the British government. All were capped by the advent of independence in 1947, but tempered by the partition of India into two states: India and Pakistan.

Vital to India's self-image as an independent nation was its constitution, completed in 1950, which put in place a secular and democratic republic.

6.2.2.1 Free Summarizer

In the Indian Ocean, India is in the vicinity of Sri Lanka and the Maldives; in addition, India's Andaman and Nicobar Islands share a maritime border with Thailand and Indonesia.

Gradually annexed by and brought under the administration of the British East India Company from the early 18th century and administered directly by the United Kingdom from the mid-19th century, India became an independent nation in 1947 after a struggle for independence that was marked by non-violent resistance led by Mahatma Gandhi.

India is a federal constitutional republic governed under a parliamentary system consisting of 28 states and 7 union territories.

The relative peace maintained by the empire during much of the 17th century was a factor in India's economic expansion, resulting in greater patronage of painting, literary forms, textiles, and architecture. Newly coherent social groups in northern and western India, such as the Marathas, the Rajputs, and the Sikhs, gained military and governing ambitions during Mughal rule, which, through collaboration or adversity, gave them both recognition and military

experience. Expanding commerce during Mughal rule gave rise to new Indian commercial and political elites along the coasts of southern and eastern India.

By the early 18th century, with the lines between commercial and political dominance being increasingly blurred, a number of European trading companies, including the English East India Company, had established coastal outposts.

The East India Company's control of the seas, greater resources, and more advanced military training and technology led it to increasingly flex its military muscle and caused it to become attractive to a portion of the Indian elite; both these factors were crucial in allowing the Company to gain control over the Bengal region by 1765 and sideline the other European companies.Its further access to the riches of Bengal and the subsequent increased strength and size of its army enabled it to annex or subdue most of India by the 1820s.India was then no longer exporting manufactured goods as it long had, but was instead supplying the British empire with raw materials, and many historians consider this to be the onset of India's colonial period.

The British Indian Empire, from the 1909 edition of The Imperial Gazetteer of India.

Fed by diverse resentments and perceptions, including invasive British-style social reforms, harsh land taxes, and summary treatment of some rich landowners and princes, the rebellion rocked many regions of northern and central India and shook the foundations of Company rule.Although the rebellion was suppressed by 1858, it led to the dissolution of the East India Company and to the direct administration of India by the British government.

Proclaiming a unitary state and a gradual but limited British-style parliamentary system, the new rulers also protected princes and landed gentry as a feudal safeguard against future unrest.In the decades following, public life gradually emerged all over India, leading eventually to the founding of the Indian National Congress in 1885.

All were capped by the advent of independence in 1947, but tempered by the partition of India into two states: India and Pakistan.

Yet, India has also been weighed down by seemingly unyielding poverty, both rural and urban; by religious and caste-related violence; by Maoist-inspired Naxalite insurgencies; and by separatism in Jammu and Kashmir and in Northeast India.

India's sustained democratic freedoms are unique among the world's new nations; however, in spite of its recent economic successes, freedom from want for its disadvantaged population remains a goal yet to be achieved.

India comprises the bulk of the Indian subcontinent, lying atop the Indian tectonic plate, and part of the Indo-Australian Plate.

India's defining geological processes began 75 million years ago when the Indian plate, then part of the southern supercontinent Gondwana, began a north-eastward driftcaused by seafloor spreading to its south-west, and later, south and south-east.

Coastal features include the marshy Rann of Kutch of western India and the alluvial Sundarbans delta of eastern India; the latter is shared with Bangladesh.

India has two archipelagos: the Lakshadweep, coral atolls off India's south-western coast; and the Andaman and Nicobar Islands, a volcanic chain in the Andaman Sea.

India is a federation with a parliamentary system governed under the Constitution of India, which serves as the country's supreme legal document.

The Constitution of India, which came into effect on 26 January 1950, states in its preamble that India is a sovereign, socialist, secular, democratic republic.

India's form of government, traditionally described as "quasi-federal" with a strong centre and weak states, has grown increasingly federal since the late 1990s as a result of political, economic, and social changes.

India is a federation composed of 28 states and 7 union territories.All states, as well as the union territories of Puducherry and the National Capital Territory of Delhi, have elected legislatures and governments, both patterned on the Westminster model.

6.2.2.1 Auto Summarizer

In the Indian Ocean, India is in the vicinity of Sri Lanka and the Maldives; in addition, Indias Andaman and Nicobar Islands share a maritime border with Thailand and Indonesia.

Gradually annexed by and brought under the administration of the British East India Company from the early 18th century and administered directly by the United Kingdom from the mid-19th century, India became an independent nation in 1947 after a struggle for independence that was marked by non-violent resistance led by Mahatma Gandhi.

The relative peace maintained by the empire during much of the 17th century was a factor in Indias economic expansion, resulting in greater patronage of painting, literary forms, textiles, and architecture. Newly coherent social groups in northern and western India, such as the Marathas, the Rajputs, and the Sikhs, gained military and governing ambitions during Mughal rule, which, through collaboration or adversity, gave them both recognition and military experience. Expanding commerce during Mughal rule gave rise to new Indian commercial and political elites along the coasts of southern and eastern India.

The East India Companys control of the seas, greater resources, and more advanced military training and technology led it to increasingly flex its military muscle and caused it to become attractive to a portion of the Indian elite; both these factors were crucial in allowing the Company to gain control over the Bengal region by 1765 and sideline the other European companies. Its further access to the riches of Bengal and the subsequent increased strength and size of its army enabled it to annex or subdue most of India by the 1820s.India was then no longer exporting manufactured goods as it long had, but was instead supplying the British empire with raw materials, and many historians consider this to be the onset of Indias colonial period.

Fed by diverse resentments and perceptions, including invasive British-style social reforms, harsh land taxes, and summary treatment of some rich landowners and princes, the rebellion rocked many regions of northern and central India and shook the foundations of Company rule. Although the rebellion was suppressed by 1858, it led to the dissolution of the East India Company and to the direct administration of India by the British government.

6.2.3 Results

	Our system	**Text Compactor**	**Free Summarizer**	**Auto Summarizer**
Algorithm used	Lexical Chain	Word Frequency	Extraction-based	Not mentioned
Time required	12 seconds	5 seconds	10 seconds	16 seconds
Output	successive sentences are continuous and related	successive sentences are not related	successive sentences are continuous and related	successive sentences are continuous and related
Start	meaningful & natural start	abrupt start	meaningful start	meaningful start
End	meaningful & natural end	abrupt end	meaningful end	meaningful end
Error caused by line breaks for next line	No	Yes	Yes	Yes
Text to speech	Yes	No	No	No
Formats of inputs taken	.txt, .docx, .doc, .pdf	.txt	.txt	.txt
Integration with external system	Can be integrated with any Java application.	Can be integrated with any Linux application	No	No

Table 6.1: Output Comparison Chart

CHAPTER 7

FUTURE SCOPE AND CONCLUSION

7. Future Scope and Conclusion

We wish to pursue further research in the following directions:

7.1 Lexical chaining algorithm:

Our method to compute lexical chains includes the gloss relations. These relations were based on the presence of gloss concept or synonym of the gloss concept in the text. We would like to pursue further research into the methods to compute the semantic similarity based on the overlap of the gloss concepts.

Lexical chains are evaluated based on their performance in identifying the sense of the word in given context. It has been proved that concepts present in the gloss of a word play an important role in the determination of the word sense. We would like to compare our system performance in this aspect with respect to other lexical chaining methods.

7.2 Document clustering:

Document clustering is a key step towards the identification of various themes in a multi-document collection. Good similarity measure plays an important role in determining the overall efficiency of the clusters. We compute the similarity measure based on the overlap of nouns (used in same sense) between two segments. Based on the study that verbs play an important role in the "action" performed in the text, we would like to investigate new methods to include the verb relations into the computation of the similarity measure.

7.3 Multi-document summarization:

Multi-document summarization is still a complex and challenging task. One problem is to find method to extract sentences to compose a coherent summary. We would like to further investigate into this problem to implement an efficient method to extract sentences from each cluster. We would like to use our sentence reduction techniques to eliminate certain portions of the extracted sentences, so as to include more content at the given compression rate.

7.4 Conclusion:

In this report, we presented a method to compute lexical chains as an efficient intermediate representation of the document. Along with normal Word Net relations, our method also included additional relations such as proper noun repetition and gloss relations in the computation of lexical chains. We identified these additional relations using semantically enhanced tool, extended Word Net. The method to include gloss relations contributes towards the better understanding of the text and enhances text coherence.

Lexical chains, until now, were mainly used to generate single document summaries. Lexical chains help identify the themes, by clustering the document collection. Indicative multi-document summaries can then be generated by selecting clusters relevant to the user's criteria and extracting sentences from each cluster. We performed intrinsic evaluation to determine the quality of the summaries generated by our approaches. We found that our system achieved better results in headline generation and in query based summarization in context.

CHAPTER 8

REFERENCES

8. REFERENCES

8.1IEEE Papers and Other Papers:

[1] Maheedhar Kolla (B.Tech), *Automatic Text Summarization Using Lexical Chains: Algorithms and Experiments* , Jawaharlal Nehru Technological University, 2002

[2] Graeme Hirst and Alexander Budanitsky *Lexical Chains* University of Toronto, August 2001, Ias¸i, Romania

8.2 Web References:

I. http://www-nlp.stanford.edu
II. http://scholar.google.co.in/scholar?q=text+summarization+algorithms
III. http://www.cse.aucegypt.edu/~rafea/CSCE590/Fall08/Shaalan/ATS_Presentation.pdf

APPENDIX

APPENDIX - I

Lexical Chains

Sample Document:

Hurricane Gilbert swept toward the Dominican Republic Sunday, and the Civil Defence alerted its heavily populated south coast to prepare for high winds, heavy rains and high seas. The storm was approaching from the southeast with sustained winds of 75 mph gusting to 92 mph. "There is no need for alarm," Civil Defence Director Eugenio Cabral said in a television alert shortly before midnight Saturday. Cabral said residents of the province of Barahona should closely follow Gilbert's movement. An estimated 100,000 people live in the province, including 70,000 in the city of Barahona, about 125 miles west of Santo Domingo. Tropical Storm Gilbert formed in the eastern Caribbean and strengthened into a hurricane Saturday night. The National Hurricane Center in Miami reported its position at 2 a.m. Sunday at latitude 16.1 north, longitude 67.5 west, about 140 miles south of Ponce, Puerto Rico, and 200 miles southeast of Santo Domingo. The National Weather Service in San Juan, Puerto Rico, said Gilbert was moving westward at 15 mph with a "broad area of cloudiness and heavy weather" rotating around the center of the storm. The weather service issued a flash flood watch for Puerto Rico and the Virgin Islands until at least 6 p.m. Sunday. Strong winds associated with the Gilbert brought coastal flooding, strong southeast winds and up to 12 feet feet to Puerto Rico's south coast. There were no reports of casualties. San Juan, on the north coast, had heavy rains and gusts Saturday, but they subsided during the night. On Saturday, Hurricane Florence was downgraded to a tropical storm and its remnants pushed inland from the U.S. Gulf Coast. Residents returned home, happy to find little damage from 80 mph winds and sheets of rain. Florence, the sixth named storm of the 1988 Atlantic storm season, was the second hurricane. The first, Debby, reached minimal hurricane strength briefly before hitting the Mexican coast last month.

Structure of chain

[score] $word_1$(sentence no.) – $word_2$(sentence no.) -- … -- $word_n$(sentence no.)

Lexical chains computed for the above text are:

[70] Hurricane(1)--winds(1)--rains(1)--storm(2)--winds(2)--Storm(6)--hurricane(6)--Hurricane(7)--storm(12)--winds(16)--winds(16)--rains(18)--Hurricane(19)--storm(19)--winds(22)--storm(23)--storm(23)--hurricane(23)--hurricane(24)

[56] Gilbert(1)--residents(4)--Gilbert(4)--miles(5)--Gilbert(6)--night(6)--a.(7)--miles(11)--miles(11)--Service(12)--Gilbert(12)--service(13)--Islands(13)--Gilbert(16)--feet(16)--feet(16)--night(18)--Residents(22)--home(22)--strength(24)--month(24)

[6] Dominican(1)--Republic(1)--province(4)--province(5)--city(5)--west(11)

[49] Sunday(1)--Defense(1)--seas(1)--Defense(3)--Director(3)--midnight(3)--Saturday(3)--movement(4)--Saturday(6)--Center(7)--Miami(7)--position(7)--Sunday(9)--north(10)--

Weather(12)--area(12)--weather(12)--center(12)--weather(13)--watch(13)--Virgin(13)--p.(13)--Sunday(15)--gusts(18)--Saturday(18)--Saturday(19)--remnants(19)--U.(19)--Gulf(21)--damage(22)--sheets(22)--season(23)

[10] Civil(1)--Civil(3)

[43] coast(1)--coast(16)--coast(18)--Coast(21)--Atlantic(23)--coast(24)

[22] southeast(2)--Caribbean(6)--latitude(9)--southeast(11)--southeast(16)

[30] mph(2)--mph(2)--mph(12)--mph(22)

[0] need(3)

[11] alarm(3)--alert(3)--flash(13)

[0] Eugenio(3)

[10] Cabral(3)--Cabral(4)

[1] television(3)--casualties(17)

[10] Barahona(4)--Barahona(5)

[0] An(5)

[1] people(5)--flood(13)

[10] Santo(5)--Santo(11)

[10] Domingo(5)--Domingo(11)

[10] National(7)--National(12)

[0] longitude(10)

[0] Ponce(11)

[30] Puerto(11)--Puerto(12)--Puerto(13)--Puerto(16)

[30] Rico(11)--Rico(12)--Rico(13)--Rico(16)

[10] San(12)--San(18)

[10] Juan(12)--Juan(18)

[0] cloudiness(12)

[0] flooding(16)

[2] reports(17)--Florence(19)--Florence(23)

[0] rain(22)

[0] Debby(24)

[0] briefly(24)

APPENDIX – II

WordNet dictionary

noun index

aachen n 1 2 @ #p 1 0 08769439

aaland_islands n 1 2 @ #p 1 0 08780510

aalborg n 1 2 @ #p 1 0 08762243

aalii n 1 2 @ #m 1 0 12740967

aalst n 1 1 @ 1 0 08850663

aalto n 1 1 @ 1 0 10806693

aar n 1 2 @ #p 1 0 09186064

aardvark n 1 2 @ #m 1 0 02082791

aardwolf n 1 2 @ #m 1 0 02118176

aare n 1 2 @ #p 1 0 09186064

aare_river n 1 2 @ #p 1 0 09186064

aarhus n 1 2 @ #p 1 0 08762104

aaron n 2 2 @ ; 2 0 10807016 10806841

aaron's_rod n 1 1 @ 1 0 12889713

aaron_burr n 1 1 @ 1 0 10874162

aaron_copland n 1 1 @ 1 0 10909929

aaron_montgomery_ward n 1 1 @ 1 0 11373897

aarp n 1 1 @ 1 0 08487149

aas n 1 1 @ 1 0 06698031

aave n 1 2 @ - 1 0 06947658

ab n 4 3 @ ~ #p 4 1 06698640 15216563 05557339 05401096

noun data

00001740 03 n 01 entity 0 003 ~ 00001930 n 0000 ~ 00002137 n 0000 ~ 04424418 n 0000 | that which is perceived or known or inferred to have its own distinct existence (living or nonliving)

00001930 03 n 01 physical_entity 0 007 @ 00001740 n 0000 ~ 00002452 n 0000 ~ 00002684 n 0000 ~ 00007347 n 0000 ~ 00020827 n 0000 ~ 00029677 n 0000 ~ 14580597 n 0000 | an entity that has physical existence

00002137 03 n 02 abstraction 0 abstract_entity 0 010 @ 00001740 n 0000 + 00692329 v 0101 ~ 00023100 n 0000 ~ 00024264 n 0000 ~ 00031264 n 0000 ~ 00031921 n 0000 ~ 00033020 n 0000 ~ 00033615 n 0000 ~ 05810143 n 0000 ~ 07999699 n 0000 | a general concept formed by extracting common features from specific examples

00002452 03 n 01 thing 0 009 @ 00001930 n 0000 ~ 04347225 n 0000 ~ 09225146 n 0000 ~ 09312645 n 0000 ~ 09367203 n 0000 ~ 09385911 n 0000 ~ 09407867 n 0000 ~ 09465459 n 0000 ~ 09468959 n 0000 | a separate and self-contained entity

0000 ~ 09474162 n 0000 ~ 09477037 n 0000 | a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects"

00003553 03 n 02 whole 0 unit 0 015 @ 00002684 n 0000 + 01462005 v 0204 + 00367685 v 0201 + 01385458 v 0201 + 00368109 v 0201 + 00784215 a 0103 ~ 00003993 n 0000 ~ 00004258 n 0000 ~ 00019128 n 0000 ~ 00021939 n 0000 ~ 02749953 n 0000 ~ 03588414 n 0000 %p 03892891 n 0000 %p 04164989 n 0000 ~ 04353803 n 0000 | an assemblage of parts that is regarded as a single entity; "how big is that part compared to the whole?"; "the team is a unit"