Sentence Compressor

by

Nur Nabilah binti Zainal Abidin

Dissertation submitted in partial fulfilment of
the requirements for the
Bachelor of Technology (Hons)
(Information and Communication Technology)

JANUARY 2014

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Redzuan

CERTIFICATION OF APPROVAL

Sentence Compressor

by

Nur Nabilah binti Zainal Abidin

A project dissertation submitted to the
Information and Communication Technology Programme
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
BACHELOR OF TECHNOLLOGY (Hons)
(INFORMATION AND COMMUNICATION TECHNOLOGY)

Approved by,

____

(MRS. FOONG OI MEAN)

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

January 2014

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or person.

NUR NABILAH BINTI ZAINAL ABIDIN

ABSTRACT

Nowadays, internet becomes the main source of information. Most people rely on the internet to find the information for research and assignments. People will try to find the right articles, journals, or web pages that are related to their task. In order to choose the right materials, they have to go through every articles, journals, and web pages to find the important points. However, it is very time-consuming to find go through every long articles. This information explosion as led to a constant state of information overload problem. As the solution, a desktop application named Sentence Compressor is developed to compress the long articles. This project aims to develop a desktop application that shortens the length of the long sentences without changing the original meaning. Integer Linear Programming (ILP) techniques is used to solve the sentence compression problem. Bilingual Evaluation Understudy (BLEU) is used to measure the quality of the produced output. Five articles were randomly selected for the experiment. The BLEU score for the articles compressed by Sentence Compressor and articles compressed by human is compared. The system performance evaluation is also done to measure the usefulness of this application. More than 65% of the respondents agreed that Sentence Compressor is useful in information searching.

ACKNOWLEDGEMENTS

I take this opportunity to express my profound gratitude and deep regards to my supervisor, Mrs. Foong Oi Mean for her exemplary guidance, monitoring and constant encouragement throughout this Final Year Project. The blessing, help and guidance given by her time to time shall carry me a long way in the journey of life on which I am about to embark.

I also take this opportunity to express a deep sense of gratitude to Mrs. Mazlina binti Mehat, the coordinator for Final Year Project I and Miss Ainol Rahmah Shazi binti Shaarani, the coordinator for Final Year Project II for their cordial support.

In addition, I would like to thank my fellow classmates for the valuable information and assistant provided by them. I am grateful for their cooperation and encouragement along the completion of Final Year Project.

I also would like to thank Petroliam Nasional Berhad (PETRONAS) for providing the sponsorship for me to undergo my study until the completion of Final Year Project.

Lastly, I thank almighty, my parents, sister and friends for their constant encouragement along the completion of this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATION AND NOMENCLATURES

API	Application Programming Interface
APIs	Application Programming Interfaces
BLEU	Bilingual Evaluation Understudy
CFG	Context Free Grammar
CS	Connective Strength
GA	General Algorithm
GSM	General Statistical Method
GUI	Graphical User Interface
ILP	Integer Linear Programing
JDK	Java Development Kit
KCS	K-Mixture Connective-Strength-Based
MSC	Multi-Sentence Compression
NLP	Natural Language Processing
POS	Part-Of-Speech
PSO	Particle Swarm Optimization
RSG	Rich Semantic Graph
RST	Rhetorical Structure Theory
SR	Sentence Removal
SRL	Semantic Role Labelling
WSS	Word Sentence Score

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Internet becomes the main source of research nowadays. Most people rely on the internet to find the related material for their research and assignments. There are a lot of web search engine on the internet (e.g. Google, Yahoo and Bing) that can help people to find the required materials for their tasks. There are also hundreds of academic search engines available to help people to find the journals and research papers online such as Google Scholar, Scopus and Web of Knowledge.

The statistic shows that 40% from the 55% Americans that used internet every day is using the search engine [1]. 42% of the world spends time on viewing web pages contents while the other 36% used email, internet banking and perform searches [1].

People who want to find articles for their research and assignments have to go through every article in order to select the right materials for their studies. They have to read the long sentences in the article to identify the important information.

There are a lot of article summarizers available to help people in summarizing articles. These article summarizers are focus on extractive summarization instead of abstractive summarization. Extractive summarization selects the sentence with highest score from an article that contains important keywords and compiles it as a summary without changing the sentence. Abstractive summarization summarizes the article by shortening the sentence without changing its original meaning. Extractive and abstractive summarization is a part of Natural Language Processing (NLP) field.

1.2 Problem Statements

There are millions of web pages, journals and research papers on the internet. With billions of new web pages and research papers being added every day, it is hard to find the article that people really need for their research. Going through every article and web pages is really time-consuming. Most people need only the important points from the article to decide whether that article is the material that they are searching for. It is extremely hard to extract useful information and keywords from the long web pages, journals and research papers. This information explosion has led to a constant state of information overload.

1.3 Significant of the Project

As the solution to the problem stated, a desktop application named Sentence Compressor is developed. Sentence Compressor is a desktop application that reduces the length of the long sentences without changing the original meaning. This application will evaluate which words in the sentence that should be removed. The summarized texts are more readable and reduce the effort to read a text. Article Summarizer will increase the readability of the texts.

The length of the sentence will be reduced by applying the Natural Language Processing (NLP) algorithm of Artificial Intelligence. Natural Language Processing (NLP) is related to the interactions between computers and human language which a part of the computer science and linguistics field [3]. In this project, Integer Linear Programming (ILP) approach is used.

1.4 Objectives of the Project

The objectives of this project are:

i. To apply Integer Linear Programming (ILP) approach for sentence compression from text article.

   Sentence compression framework and algorithm proposed by Clarke and Lapata [26] that used ILP approach is applied in this project to compress the sentence. This approach will reduce the length of the long sentences without changing its meaning before generating the final summary.

ii. To develop a desktop application that reduces long sentences into shorter ones without changing the original meaning.

The second objective of this project is to develop a desktop application that can reduce the long article by reducing the length of the long sentences. However, the shortened sentences will not change the original meaning of the sentence. This application will make the selecting job of the journals and research papers more convenient.

1.5 Scope of Study

In order to develop this application, there are several studies is done. The scopes of study for this project are:

i. Research on the Natural Language Processing (NLP) technique.

NLP is one of the techniques in Artificial Intelligence area. NLP field is quite broad. However, for this project, the NLP technique is narrowed down to automatic summarization. Automatic summarization is one of the NLP major tasks. The methods for automatic summarization is identified and investigated in order to decide which method should be used for this application. In order to evaluate the summarized texts, an evaluation technique must be chosen. To decide which evaluation technique should be used; a brief study on evaluation techniques of automatic summarization must be done. NLP knowledge must be mastered before developing this mobile application.

ii. Research on the Integer Linear Programming (ILP).

ILP is used in this project to determine the score for every word in every sentence and determine which word should be removed. ILP problems are actually Linear Programming (LP) problem, but the variables must be non-negative integers. The formula of ILP problems is similar as LP problems, but ILP has additional constraints where all decisions must be non-negative integers.

1.6 Relevancy of the Project

This project applies the Artificial Intelligence (AI) knowledge. Artificial intelligent scope is really broad. So, this project focuses on Natural Language Processing (NLP) which is a part of the Artificial Intelligence field.

Automated summarization is one of the NLP tasks. This project implements the automated summarization in order to achieve the objective of reducing the length of the long sentences to shorter ones.

This application will solve the problem addressed in the problem statements where this application will summarize the long article and reduce the length of the sentence.

1.7 Feasibility of the Project

1.7.1 Technical Feasibility

Sentence Compressor is a desktop application. Java programming language is used to develop this project. The developer of this project is really familiar with Java programming language.

1.7.2 Operational Feasibility

The problem on the difficulty to find the information from the long article has been addressed by a lot of researchers. The summarized text is able to increase the readability of the text. Hiring people to do the summarizing task is costly. With this automated summarization application, the effort to manually summarize the article can be reduced. This project is operationally feasible because it can solve the problems addressed in the Problem Statements section.

1.7.3 Schedule Feasibility

The deadline of this project can be met because the proposed framework and algorithm is applied in this project. Hence, Final Year Project II is focused on the development of the sentence compressor. So, the development of this project can be done before the end of Final Year Project II.

1.8 Research Gap

There are a lot of research, studies and software for text summarization. Most of the text summarizer software developed is using the extraction based method. Extraction based method is only select and combine the most important sentences as the summary. This method takes the whole sentences with the highest score and combines the sentences as the summary without paraphrasing it.

Most of the software and research are using extraction based method because it is easy to adapt to larger sources. However, because this method only limited to the extraction of passage, sentences, or phrases, the resulting summaries may be incoherent [12]. The extractive summary is only rely on extracting the sentences and only concern about the content of the summary [13]. Most of the automatic summary also hardly similar to the manually summarized article by human. Microsoft Word Summarizer is only 39% similar to the manually generated summaries. [14].

In this project, sentence compression approach is applied to reduce the length of the long sentences without changing the original meaning. In this sentence compression approach, the unimportant words will be pruned or removed.

CHAPTER 2

LITERATURE REVIEW AND THEORY

2.1 Summarized Texts and Readability

Most people have problems to assimilate long and complex texts. This problem does not only occur to those who have visual impairments or dyslexia, but also to those who have different mother tongue. People who are in need to find a quick summary of an article, journal or research papers also face the same problem.

In a study by Smith and Jönsson, it is proven that the summarized texts are more readable, indicating that summarization can be used to reduce the effort to read a text [4]. The investigation was done by using computer software named CogSum; a summarizer based on word space model random indexing (RI) and PageRank. The evaluations on readability were done using a number of automatic measures of readability [4].

After the evaluations, it was concluded that the automatic summarization gave positive impact on readability for texts [4]. So, this shows that summarization can make a text easier to read. The summarized content and important sentences are more helpful than the whole article to determine the right articles for assignments and research.

The aim of this project is to increase the readability of web pages, journals and research paper that are posted online by reducing the length of the sentence. This can be done by using the Natural Language Processing (NLP) algorithm. The users of this application should feel that the summarized text is easier to read and save their time in searching for the information.

2.2 Natural Language Processing

(NLP) is related to the interactions between computers and human language which a part of the computer science and linguistics field [3]. This technique makes machine become more human. There are many applications developed that use NLP.

Mainly, the application is divided into two parts which are dialogue-based applications and text-based applications. Text-based applications are used to search for a certain topic or a keyword in a database, extract information from large documents translating one language to another language or summarizing for different purposes. Dialogue based applications can provide automatic answers for answering systems over a telephone without an operator, teaching systems, voice controlled machines and general problem solving. The examples of dialogue based applications are Cleverbot, SimSimi and Siri.

Sentence Compressor is a text-based application since it is functioning as the sentence reducer. There are a lot of major tasks in NLP such as machine translation, text-to-speech, automatic summarization and text simplification. This project is based on automatic summarization and text simplification.

Natural Language Processing (NLP) plays important role in machine-human interaction [3]. Hiring a human to summarize a long article will be very costly especially hiring those with high qualification. By using NLP, the cost on hiring a human to manually summarize an article can be avoided.

NLP has some limitations that require improvements. The language spoken by human is highly ambiguous and makes it difficult to understand. People over the world speak different languages. It is hard for a system to be 100% accurate. So, the Sentence Compressor will focus on English language since it is the main intermediate language in the world.

2.3 Sentence Compression

Automatic Summarization helps in extracting the information, keywords, and the contents so that the user will be presented by the most important part of the content [6]. The interest on automatic summarization is increasing as the quantity of the data has increased. There are four methods of summarization which are; abstraction-based summarization, extraction-based summarization, aided summarization and maximum entropy-based summarization.

However a common distinction is between extraction-based summarization versus abstraction-based summarization. Extraction-based summarization extracts the most important sentences from the original text [4]. Abstraction-based summarization breaks down the text and rebuild as a complete rewrite to convey a general idea of the original text [4].

The extraction-based summarization selects an important individual word or phrases, then select the whole sentence to create a short paragraph summary. So, if this approach is being used, the long article will be shortened by just displaying some of the full sentences from the article that are considered as important.

The abstraction-based summarization paraphrases a long sentence to be shorter. It creates an abstract synopsis of an article like it is summarized by a human. Researchers believe that abstraction-based summarization can provide shorter summary compare to extraction-based summarizations. However, the programs that use abstraction-based summarization are harder to develop. This project is focused on abstraction-based summarization in order to achieve the objective; shorten the length of the long sentences.

Sentence compression recently attracted a lot of researchers' attention. Commonly, sentence compression is expressed as a word deletion problem. The aim of sentence compression is to remove some words to produce a new sentence with targeted compression [28]. The sentence compression problem had been studies in supervised and unsupervised modelling paradigms.

The first sentence compression problem is tackled by [29].She used multiple knowledge sources to determine which phrases in a sentence to remove. The important part of her system is the grammar checking module that specifies which sentential constituents are grammatically obligatory and should therefore be present in the compression. A phrase is removed only if it is not grammatically obligatory, not the focus of the local context and has a reasonable deletion probability which is estimated from the parallel corpus.

A lot of research on sentence compression relies on the corpus data for modelling the compression. Quite a huge number of approaches for sentence compression are based on noisy-channel model [28]. These approaches consist of a language model which is used to guarantee that compression output is grammatically correct, a channel model to capture the probability that the source sentence is an expansion of target compression, and a decoder which is used to search for the compression that maximize the language model and channel model.

Sentences are represented by a rich feature space in discriminative models and the goal is to learn which words should be deleted in a sentence [28]. For example, in [28], they used decision-tree model where the compression is performed deterministically through a tree rewriting process.

Meanwhile, [30] proposed a model that automatically transcribed spoken text instead of written text. Their model generates compression through word deletion without using parallel data or syntactic information in any way. This approach searches for the compression with the highest score by using dynamic programming algorithm. to ensure the compression is grammatically correct, language model is used in scoring function.

Existing approaches do not model global properties of the compression problem for a good reason. Finding the best compression for every sentence by listing down all possible compression can become intractable for too many constraints and overly long sentences [26]. Usually dynamic programming is used to solve the decoding problem efficiently [31].

The approach proposed by [26] for sentence compression that considers constraints on the compression output allows the optimal solution to be found. So, instead of using dynamic programming for decoding the best solution, [26] used Integer Linear Programming [ILP] technique. ILP is used to make a final decision that is consistent with the constraints and likely according to the classifier.

CHAPTER 3

METHODOLOGY

3.1 Introduction to Linear Programming and Integer Linear Programming

This project uses the system framework proposed by Clarke and Lapata [26]. This framework is using Integer Linear Programming (ILP) approach to solve the decision of word deletion. Section 3.1.1 and 3.1.2 is the brief introduction of Linear Programming (LP) and Integer Linear Programming (ILP).

3.1.1 Linear Programming (LP)

Linear Programming (LP) problems are optimization problems with constraints. LP consists of three important keywords [26]:

- Decision Variables: The decision variables can be controlled to assign the optimal values.
- A linear Function (the Objective Function): The value of decision variables assigned influence this linear function.
- Constraints: This is the restriction where the problems will only allow the decision variables to take certain values.

The example by Winston and Venkataramanan [27] is used to explain these three keywords. Telfa Corporation is the manufacturer for tables and chairs. 1 hour labour and 9 square board feet of wood is required to produce a table; and 1 hour labour and 5 square board feet of wood is needed to produce a chair. There is 45 square board feet of woods and 6 hours labour available to produce both; the chairs and the tables. In this case, Telfa should determine the number of tables and chairs that must be produced in order to maximize the profit. Let say, a table gives 8 GBP profit and chair gives 5 GBP.

The first step, the decision variable must be determined. In this case:

x_1 = Number of tables manufactured

x_2 = Number of chairs manufactured

The linear function or objective function is:

$$\text{Profit} = 8x_1 + 5x_2$$

The constraints in this situation are:

Labour constraint	:	$x_1 + x_2 \leq 6$
Wood constraint	:	$9x_1 + 5x_2 \leq 45$
Variable constraints	:	$x_1 \leq 0$
		$x_2 \leq 0$

The production for tables and chairs must not more than 6 labour hours, and must not use more than 45 square board feet of wood. The amount of chair and table must be more than zero because it is impossible to create negative number of tables and chairs.

To calculate the maximum profit:

$$x_1 = 6 - x_2$$

$$x_2 = 6 - x_1$$

$$9\,(6 - x_2) + 5x_2 = 45$$

$$54 - 9x_2 + 5x_2 = 45$$

$$x_2 = 9/4 = 2.25$$

$$x_1 = 15/4 = 3.75$$

$$\text{Profit} = 8\,(3.75) + 5\,(2.25) = 41.25 \text{ GBP}$$

From the above calculation, to maximize the profit of 41.25 GBP, Telfa should produce 3.75 tables and 2.25 tables. However, this is impossible since the company cannot expect people to buy fractions of tables and chairs. To solve this issue, Integer Linear Programming (ILP) is used.

3.1.2 Integer Linear Programming (ILP)

Integer Linear Programming (ILP) problems are LP problems in which some or all of the variables are required to be non-negative integers. It applies the similar manner as LP problems, but with added constraints where all decision variables must take non-negative integer values. In this project, the ILP problems only take arbitrary values that are restricted to 0 and 1. So, instead of giving the non-integer value, ILP gives the integer value that helps to solve the example problem that is previously stated. The number of tables and chairs that need to be manufactured in order to maximize the profit will be in the whole number instead of fractions.

3.2 Sentence Compressor Framework

This application is using the framework and algorithm proposed by Clarke and Lapata [26]. However, the system framework diagram in Figure 3.1 is drawn based on the research paper written by them. They did not come out with specific system framework diagram for the sentence compression using ILP. Sentence Compressor framework is divided into four phases; keywords extraction, calculate objective function, add constraints and compress sentence. Figure 3.1 shows the system framework used for Sentence Compressor.

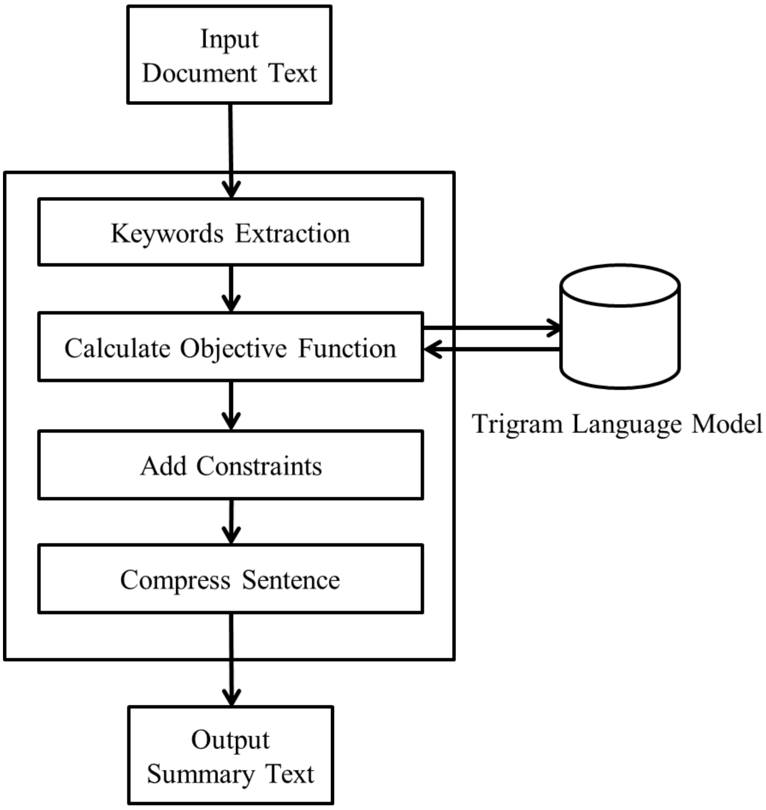


FIGURE 3.1: Sentence Compressor System Framework

Each of the phases is explained in the following subsections.

3.2.1 Keywords Extraction

In this phase, the system will evaluate each sentence in the article. If the sentence contains unimportant word, the clause that contains that unimportant word will be deleted. The sentence that does not contain unimportant word will remain as it is. The frequency of occurrence for each word will be calculated to determine whether the word is important or not.

3.2.2 Calculate Objective Function

The first process to calculate the objective function is creating some additional decision variables. The following are decision variables to calculate the objective function.

$$\delta_i = \begin{cases} 1 & if\ x_i\ is\ in\ the\ compression \\ 0 & otherwise \end{cases} \quad \forall_i \in [1 \dots n]$$

$$\alpha_i = \begin{cases} 1 & if\ x_i\ starts\ the\ compression \\ 0 & otherwise \end{cases} \quad \forall_i \in [1 \dots n]$$

$$\beta_{ij} = \begin{cases} 1 & if\ x_i\ x_j\ ends\ the\ compression \\ 0 & otherwise \end{cases} \quad \begin{matrix} \forall_i \in [0 \dots n-1] \\ \forall_j \in [i+1 \dots n] \end{matrix}$$

$$\gamma_{ijk} = \begin{cases} 1 & if\ x_i\ x_j\ \ x_k\ \ is\ in\ the\ compression \\ 0 & otherwise \end{cases} \quad \begin{matrix} \forall_i \in [0 \dots n-2] \\ \forall_j \in [i+1 \dots n-1] \\ \forall_k \in [j+1 \dots n] \end{matrix}$$

Every δ_i , α_i , β_{ij} , γ_{ijk} will be assigned with binary value, either 1 or 0 based on the condition defined so that the objective function can be calculated in the next stage. The following is the objective function used for sentence compression decision.

$$\max z = \sum_{i=1}^{n} \alpha_i . P(x_i|start) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} \gamma_{ijk} . P(x_k|x_i x_j)$$
$$+ \sum_{i=0}^{n-1} \sum_{j=i+1}^{n} \beta_{ij}\ .\ P(end|x_i x_j)$$

This equation calculates the sum of α_i with the probability (according to the language model) of word x_i to start the compression, add to the sum of $\gamma_{ijk}$with the probability of the words to be in the compression and add to the sum of β_{ij} with the probability of the words $x_i x_j$ to end the compression. The decision variables are restricted to be binary (1 or 0).

$$\delta_i\ ,\alpha_i\ ,\beta_{ij}\ ,\gamma_{ijk} = 0\ or\ 1$$

In this phase, IBM ILOG CPLEX Optimization Studio and CPLEX library is used to make analytical decision support. The objective function is computed by using this software. In this project, the trial version of IBM ILOG CPLEX Optimization Studio is used. The trial version limits the computation up to only certain number of problems and constraints.

3.2.3 Add Constraints

In this phase, the constraints are added to the computation made in IBM ILOG CPLEX Optimization Studio. There are five constraints proposed by Clarke and Lapata [27]. These five constraints is used in this project without being changed.

Constraint 1: Exactly one word can begin a sentence.

$$\sum_{i=1}^{n} \alpha_i = 1$$

Constraint 2: If a word exists in the sentence, it must either start a sentence or be preceded by two other words or one other word and the 'start' token x_0.

$$\delta_k - \alpha_k - \sum_{i=0}^{k-2} \sum_{j=1}^{k-1} \gamma_{ijk} = 0$$

$$\forall k : k \in [1 \dots n]$$

Constraint 3: If a word exists in the sentence, it must either be preceded by one word and followed by another, or it must be preceded by one word and end the sentence.

$$\delta_j - \sum_{i=0}^{j-1} \sum_{k=j+1}^{n} \gamma_{ijk} - \sum_{i=0}^{j-1} \beta_{ij} = 0$$

$$\forall j : j \in [1 \dots n]$$

Constraint 4: If a word exist in the sentence, it must be followed by two words or followed by one word and then the end of the sentence or it must be preceded by one word and end the sentence.

$$\delta_i - \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} \gamma_{ijk} - \sum_{j=i+1}^{n} \beta_{ij} - \sum_{h=0}^{i-1} \beta_{hi} = 0$$

$$\forall i: i \in [1 \dots n]$$

Constraint 5: Exactly one word pair can end the sentence.

$$\sum_{i=0}^{n-1} \sum_{j=i+1}^{n} \beta_{ij} = 1$$

3.2.4 Compress Sentence

In this phase, the IBM ILOG CPLEX Optimization Studio will return the decision in binary value for each word. The word with 0 score will be deleted from the sentence, while the word with 1 score will remain in the sentence.

3.3 Project Method and Activities

Prototyping methodology is used throughout the development of this software. There are four main phases in this methodology which are planning, analysis, design and implementation. A software prototype is produced after performing analysis, design and implementation phases concurrently. The cycle is repeated continuously until the prototype successfully meets the requirements.

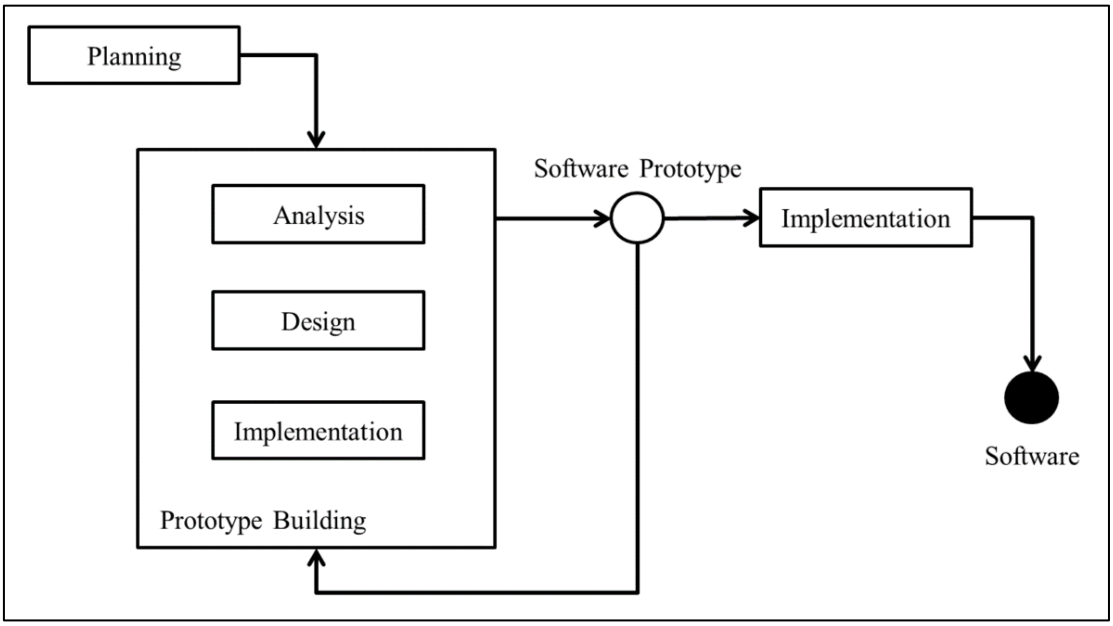


FIGURE 3.2: Prototyping Methodology

3.3.1 Project Activities

There are a few project activities for each phase. The activities are as follow:

i. Planning

   a. Define problem

      The current problems and limitations while searching for information from web pages, journals and research papers are identified.

   b. Formulate objective

      The aim and goal of the project is set.

ii. Analysis

   a. Review concept and theories

      The concepts of Natural Language Processing (NLP) and automatic summarization are analysed.

   b. Review previous research findings

      The previous research project regarding article summarizing and study on the effect of summarised article are analysed.

   c. Collect data

      Data regarding to NLP, automatic summarization and frameworks are collected.

d. Analyse data

The possible frameworks and APIs are identified and analysed.

iii. Design

a. Design application

The requirements are listed to make the design of the application.

iv. Implementation

a. Application implementation

The application is developed.

b. Testing

The application developed is tested.

3.3.2 Key Milestones

TABLE 3.1: Key Milestone

Milestone	Completion Date
List possible approach	29th October 2013
List possible tools	3rd December 2013
Choose sentence compression framework	13th December 2013
Apply sentence compressor framework	15th December 2013
System performance evaluation	27th March 2014
Compressed sentence evaluation	4th April 2014
Project delivery	8th April 2014

3.3.3 Gantt Chart

Please refer Appendix I.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Human Summary Procedure

This project aims to produce an abstractive summary by using automated text summarization. There are a lot standard procedure and guideline for human to summarize the article. This project adapts the procedures to summarize an article written in [23].

The following describes the procedures to summarize the text according to [23]:

1. A person must understand the text article before summarizing.
2. Important sentences for the summary must be underlined.
3. The text article must be divided into paragraphs.
4. Heading for different paragraphs should be found.
5. Keywords that important for the summary must be jot down.
6. Original sentence must be paraphrased.

There are a lot of structures of sentence that human need to follow. Hence, the ambiguity of human language make automated summarization task become very challenging. Human summarization usually changes the tense of the sentence. For example, statements that were made in the past can use past tense or past perfect.

Due to this ambiguity, this project will focus on reducing the original sentence buy just pruning the unimportant words. Meaning that, the finalized summary will consist of the shortened sentences without changing the tense of the sentence. For example, the original sentence is “She is walking at a peaceful and beautiful park”. In this case, the adjective words are not important. After pruning the word “peaceful” and “beautiful”, the shortened sentence will be, “She is walking at a park.”

In order to achieve the objective of shortening the sentence, this project will refer to the basic English grammar sentence construction. According to [24], there are three types of sentence:

- **Simple sentence**

  A simple sentence contains a single subject and predicate. This simple sentence should contain only a verb. The example of a simple sentence is “The woman is reading a book”.
- **Compound sentence**

  A compound sentence is made up of two or more simple sentences combined by using a conjunction such as ‘and’, ’or’, and ‘but’. So, this sentence is made up of more than one independent clause joined together by using the conjunction. Each clause can stand alone as a sentence. For example, “The woman is reading a book while the man is drinking a cup of coffee” can stand alone as “The woman is reading a book. The man is drinking a cup of coffee”. In this project, we will prune the unimportant clause by detecting the important keywords of the sentence.
- **Complex sentence**

  A complex sentence basically describes more than one idea in a sentence. This sentence usually has more than one verb and clause in it. For example, “The woman is reading a book while drinking a cup of coffee at the park near her house”. This is the most challenging part; to reduce the sentence since we have to come out with a specific rule in order to reduce this sentence without any grammatical errors.

4.2 Project Prototype

A desktop application is developed to test the algorithm used for sentence compression. The desktop application is built by using JAVA programming language. This desktop application is named as "Sentence Compressor". Sentence Compressor only allows the users to import the text file that contains the sentence that need to be reduced from any drives.

Figure 4.1 shows the main section of the Sentence Compressor. The top text area will display the extracted sentence from the text file. The bottom text area will display the compressed article. Users are not allowed to press the "Compress" button and "Information" menu button until they import a file that contains sentence to be compressed.

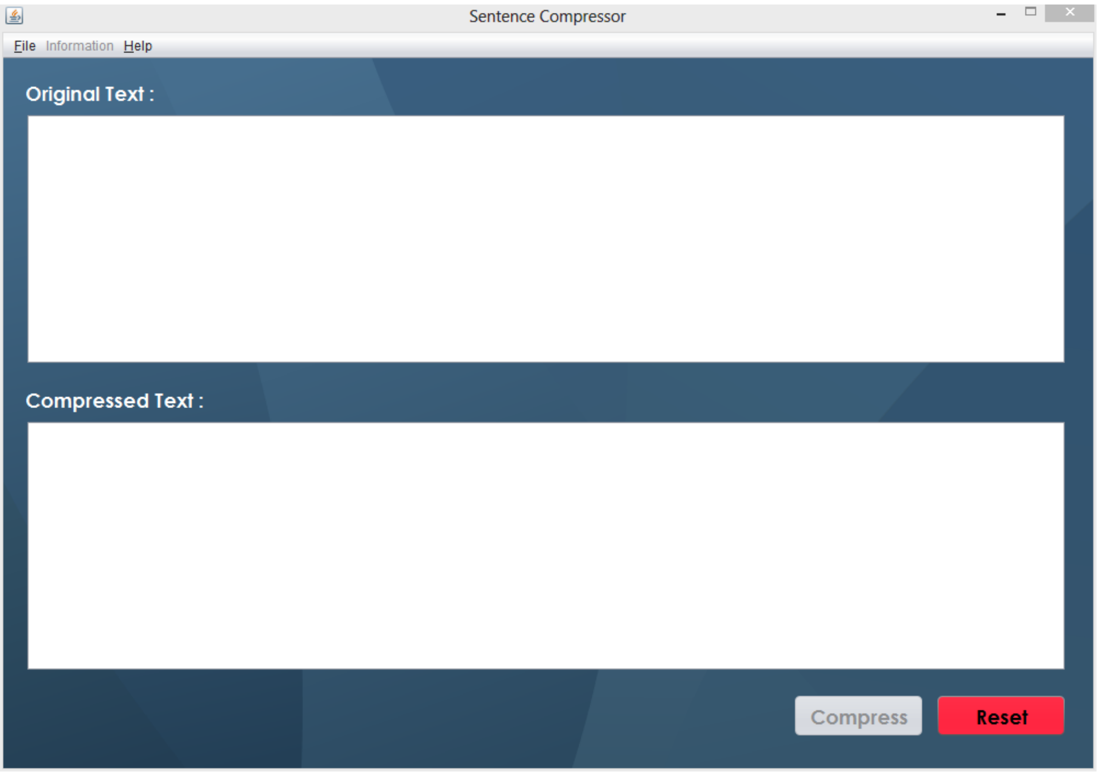


FIGURE 4.1: Main Section of Sentence Compressor

To import a text file, users have to click the "File" menu button and choose "Import File" option.

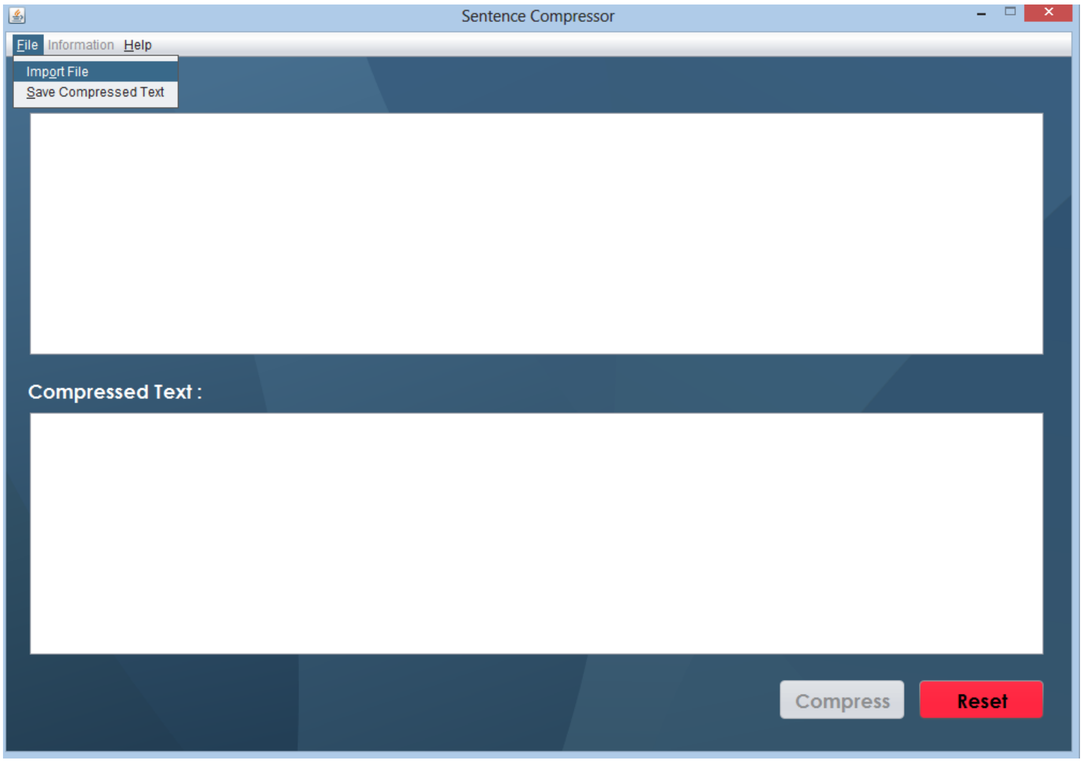
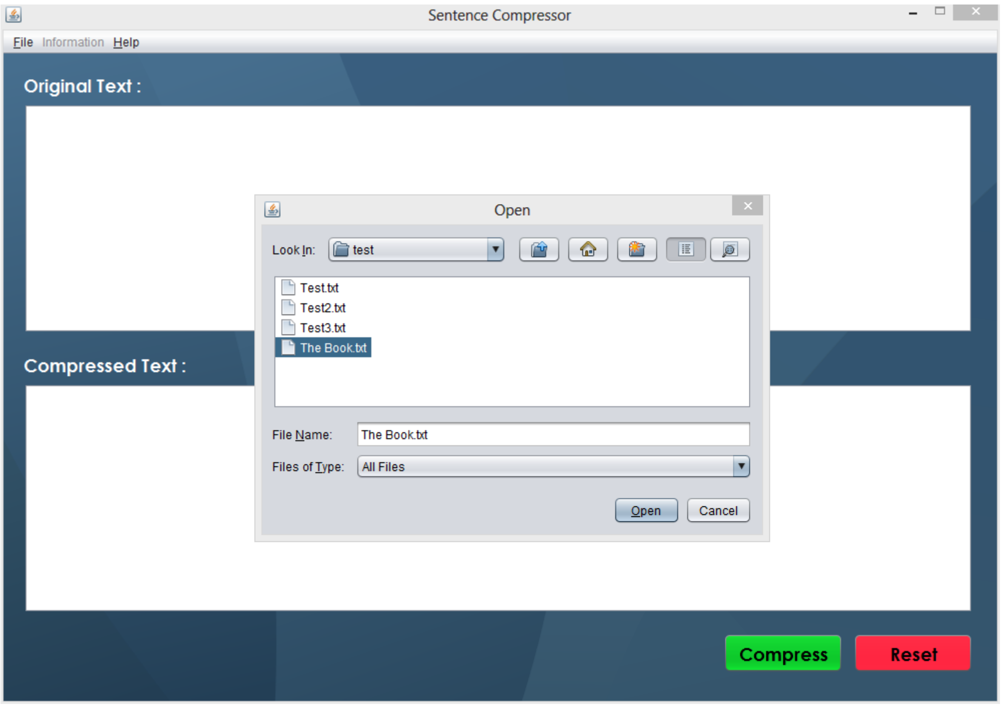


FIGURE 4.2: User Clicks Import File Option

After clicking the “Import File” option, a file chooser will be popped up to allow the user to choose a text file that contains the sentence need to be compressed from any folder.



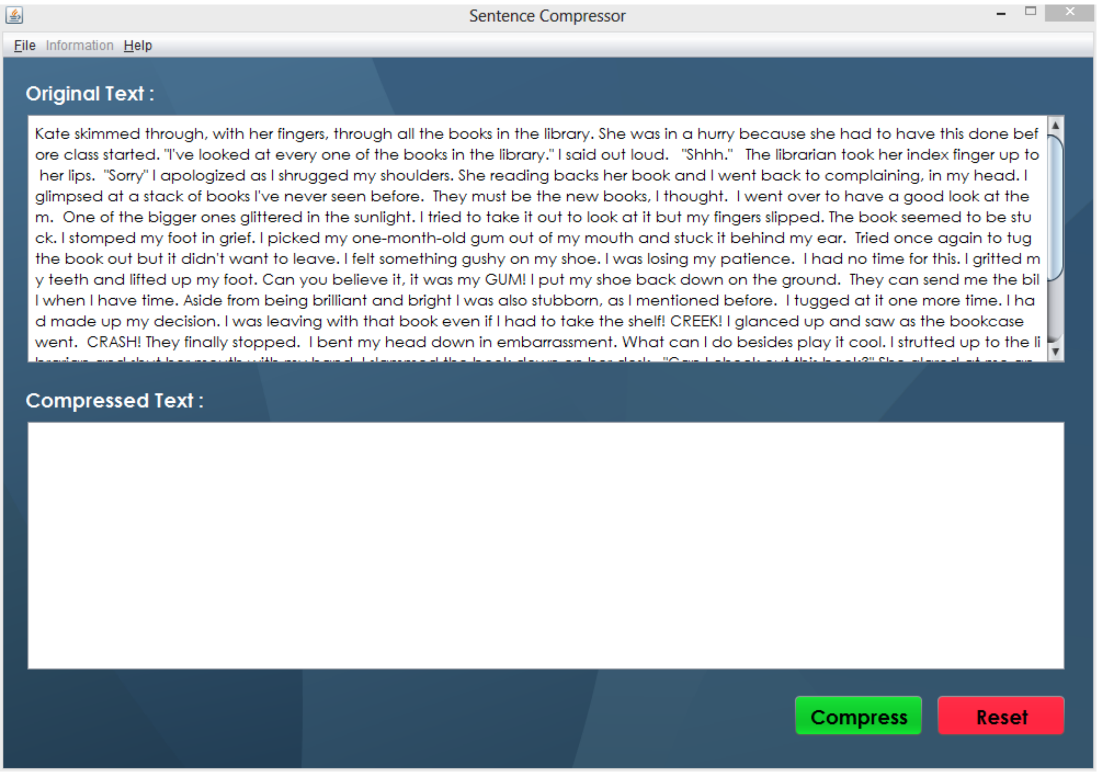
Look In: test

File Name: The Book.txt

Files of Type: All Files

FIGURE 4.3: File Chooser of Sentence Compressor

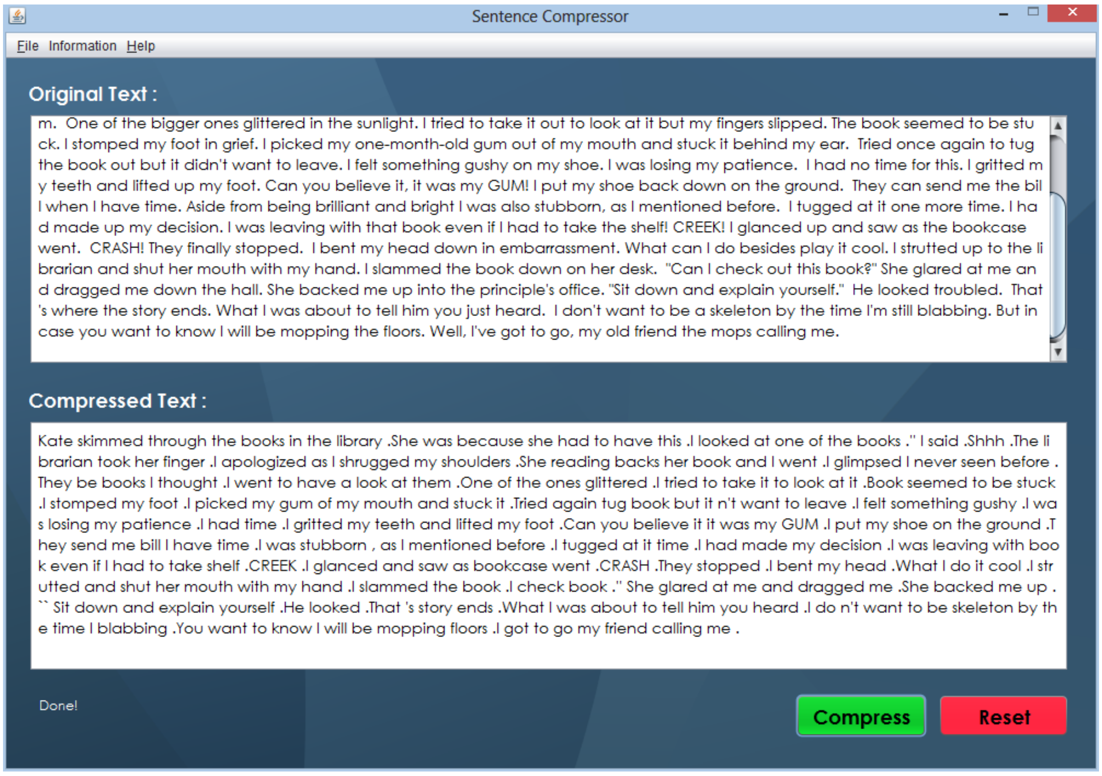
Sentence Compressor will automatically extract the sentences from the selected text file and display the content in the Original Text area column. The button "Compress" will automatically turn green after the contents of the file are extracted.



Kate skimmed through, with her fingers, through all the books in the library. She was in a hurry because she had to have this done before class started. "I've looked at every one of the books in the library." I said out loud. "Shhh." The librarian took her index finger up to her lips. "Sorry" I apologized as I shrugged my shoulders. She reading backs her book and I went back to complaining, in my head. I glimpsed at a stack of books I've never seen before. They must be the new books, I thought. I went over to have a good look at them. One of the bigger ones glittered in the sunlight. I tried to take it out to look at it but my fingers slipped. The book seemed to be stuck. I stomped my foot in grief. I picked my one-month-old gum out of my mouth and stuck it behind my ear. Tried once again to tug the book out but it didn't want to leave. I felt something gushy on my shoe. I was losing my patience. I had no time for this. I gritted my teeth and lifted up my foot. Can you believe it, it was my GUM! I put my shoe back down on the ground. They can send me the bill when I have time. Aside from being brilliant and bright I was also stubborn, as I mentioned before. I tugged at it one more time. I had made up my decision. I was leaving with that book even if I had to take the shelf! CREEK! I glanced up and saw as the bookcase went. CRASH! They finally stopped. I bent my head down in embarrassment. What can I do besides play it cool. I strutted up to the li

FIGURE 4.4: Extracted Sentence from Text File

After the user pressed the "Compress" button, this application will start compressing the sentences. The compressed sentences are displayed in the Compressed Text area column as shown in figure 4.5.

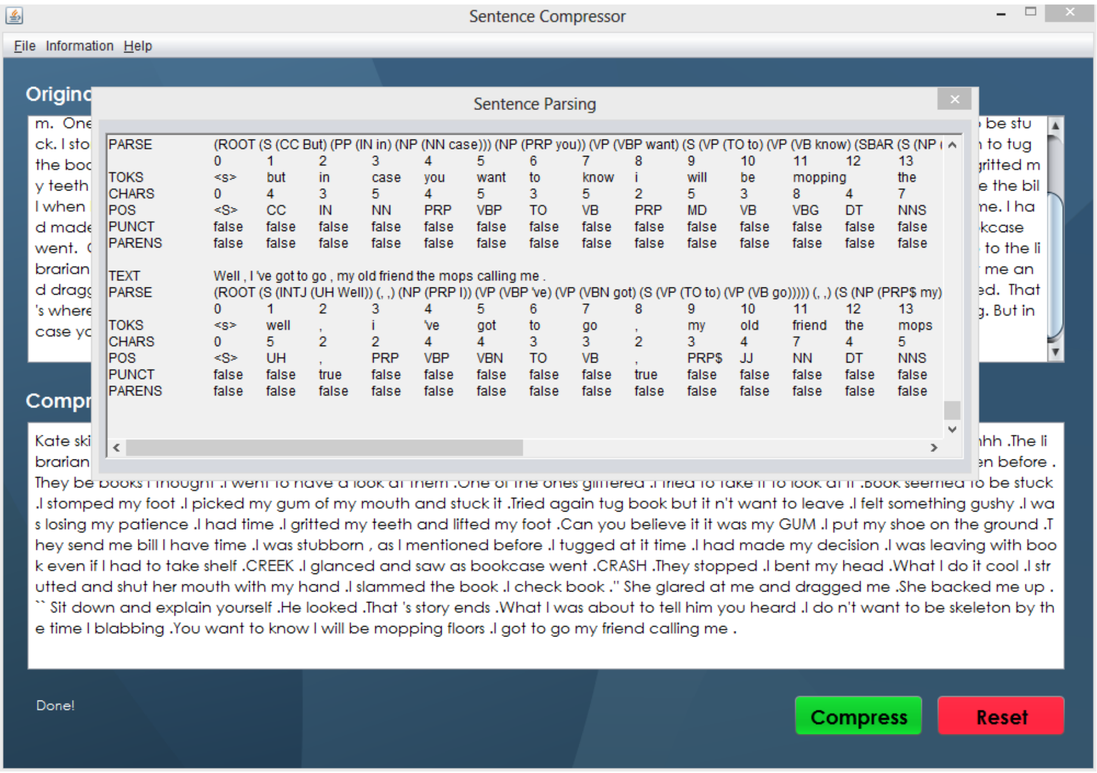


```
m.  One of the bigger ones glittered in the sunlight. I tried to take it out to look at it but my fingers slipped. The book seemed to be stu
ck. I stomped my foot in grief. I picked my one-month-old gum out of my mouth and stuck it behind my ear.  Tried once again to tug
the book out but it didn't want to leave. I felt something gushy on my shoe. I was losing my patience.  I had no time for this. I gritted m
y teeth and lifted up my foot. Can you believe it, it was my GUM! I put my shoe back down on the ground.  They can send me the bil
l when I have time. Aside from being brilliant and bright I was also stubborn, as I mentioned before.  I tugged at it one more time. I ha
d made up my decision. I was leaving with that book even if I had to take the shelf! CREEK! I glanced up and saw as the bookcase
went.  CRASH! They finally stopped.  I bent my head down in embarrassment. What can I do besides play it cool. I strutted up to the li
brarian and shut her mouth with my hand. I slammed the book down on her desk.  "Can I check out this book?" She glared at me an
d dragged me down the hall. She backed me up into the principle's office. "Sit down and explain yourself."  He looked troubled.  That
's where the story ends. What I was about to tell him you just heard.  I don't want to be a skeleton by the time I'm still blabbing. But in
case you want to know I will be mopping the floors. Well, I've got to go, my old friend the mops calling me.
```

```
Kate skimmed through the books in the library .She was because she had to have this .I looked at one of the books ." I said .Shhh .The li
brarian took her finger .I apologized as I shrugged my shoulders .She reading backs her book and I went .I glimpsed I never seen before .
They be books I thought .I went to have a look at them .One of the ones glittered .I tried to take it to look at it .Book seemed to be stuck
.I stomped my foot .I picked my gum of my mouth and stuck it .Tried again tug book but it n't want to leave .I felt something gushy .I wa
s losing my patience .I had time .I gritted my teeth and lifted my foot .Can you believe it it was my GUM .I put my shoe on the ground .T
hey send me bill I have time .I was stubborn , as I mentioned before .I tugged at it time .I had made my decision .I was leaving with boo
k even if I had to take shelf .CREEK .I glanced and saw as bookcase went .CRASH .They stopped .I bent my head .What I do it cool .I str
utted and shut her mouth with my hand .I slammed the book .I check book ." She glared at me and dragged me .She backed me up .
`` Sit down and explain yourself .He looked .That 's story ends .What I was about to tell him you heard .I do n't want to be skeleton by th
e time I blabbing .You want to know I will be mopping floors .I got to go my friend calling me .
```

FIGURE 4.5: Displayed Compressed Sentences

Users can read the information for every parsed sentence in the article by clicking the "Information" menu button. Figure 4.6 shows the parsed sentence for every sentence in the text article. "PARSE" shows the grammar dependencies for every word in the sentence. "TOKS" shows the separated tokenized word. "CHARS" shows the number of characters for every word. "POS" shows the tokenized words with specific label from Part of Speech (POS) tagger. "PUNCT" shows whether the tokenized words are punctuation or not. "PARENS" shows whether the tokenized word is inside the bracket or not.



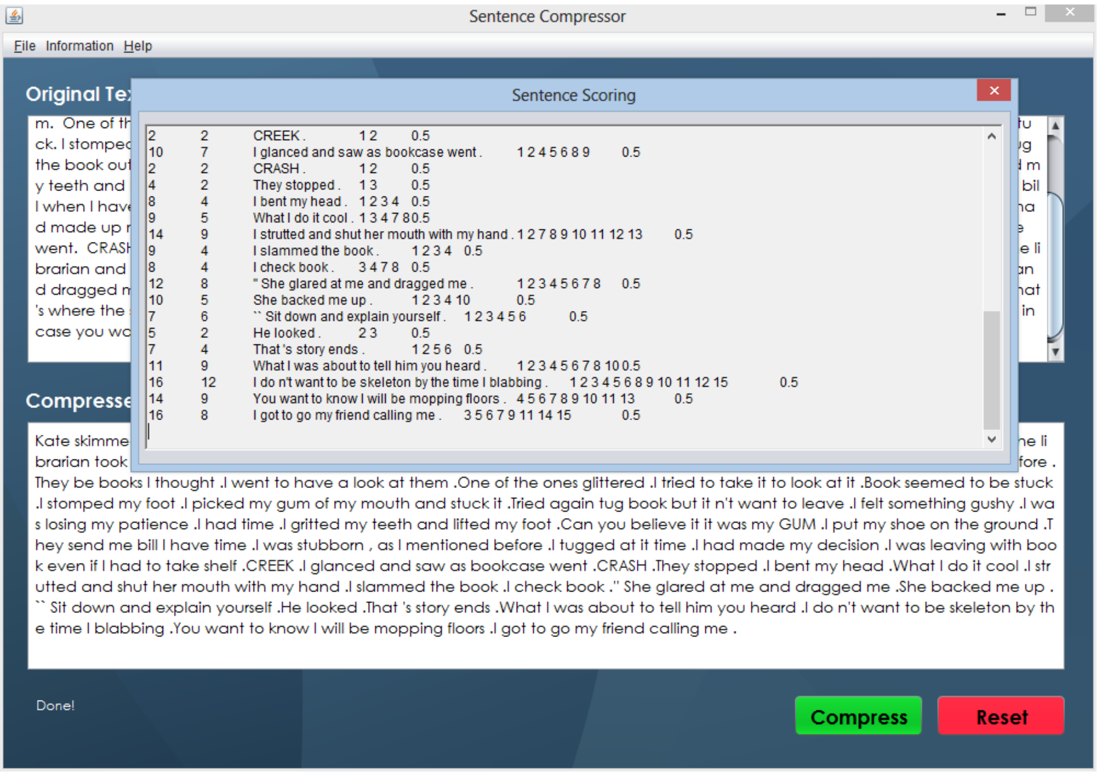
```
PARSE     (ROOT (S (CC But) (PP (IN in) (NP (NN case))) (NP (PRP you)) (VP (VBP want) (S (VP (TO to) (VP (VB know) (SBAR (S (NP
          0      1      2      3      4      5      6      7      8      9      10     11       12     13
TOKS      <s>    but    in     case   you    want   to     know   i      will   be     mopping         the
CHARS     0      4      3      5      4      5      3      5      2      5      3      8        4      7
POS       <S>    CC     IN     NN     PRP    VBP    TO     VB     PRP    MD     VB     VBG      DT     NNS
PUNCT     false  false  false  false  false  false  false  false  false  false  false  false    false  false
PARENS    false  false  false  false  false  false  false  false  false  false  false  false    false  false

TEXT      Well , I 've got to go , my old friend the mops calling me .
PARSE     (ROOT (S (INTJ (UH Well)) (, ,) (NP (PRP I)) (VP (VBP 've) (VP (VBN got) (S (VP (TO to) (VP (VB go))))) (, ,) (S (NP (PRP$ my)
          0      1      2      3      4      5      6      7      8      9      10     11       12     13
TOKS      <s>    well   ,      i      've    got    to     go     ,      my     old    friend   the    mops
CHARS     0      5      2      2      4      4      3      3      2      3      4      7        4      5
POS       <S>    UH     ,      PRP    VBP    VBN    TO     VB     ,      PRP$   JJ     NN       DT     NNS
PUNCT     false  false  true   false  false  false  false  false  true   false  false  false    false  false
PARENS    false  false  false  false  false  false  false  false  false  false  false  false    false  false
```

.I stomped my foot .I picked my gum of my mouth and stuck it .Tried again tug book but it n't want to leave .I felt something gushy .I was losing my patience .I had time .I gritted my teeth and lifted my foot .Can you believe it it was my GUM .I put my shoe on the ground .They send me bill I have time .I was stubborn , as I mentioned before .I tugged at it time .I had made my decision .I was leaving with book even if I had to take shelf .CREEK .I glanced and saw as bookcase went .CRASH .They stopped .I bent my head .What I do it cool .I strutted and shut her mouth with my hand .I slammed the book .I check book ." She glared at me and dragged me .She backed me up . `` Sit down and explain yourself .He looked .That 's story ends .What I was about to tell him you heard .I do n't want to be skeleton by the time I blabbing .You want to know I will be mopping floors .I got to go my friend calling me .

FIGURE 4.6: Sentence Parsing Window

Besides that, the user also can see the information about the compressed sentence separately with additional information. Figure 4.7 shows the Sentence Scoring Window for Sentence Compressor. The first column shows the original number of words in the sentence. The second column shows the number of words after the sentence is compressed. The sentence displayed is the compressed sentence or also called as the output. The fourth column shows the original position of the words in the original sentence. The last column shows the compression rate. In figure 4.7, it shows 0.5 which means 50% compression rate.



m. One of th
ck. I stomped
the book out
y teeth and
I when I have
d made up r
went. CRASH
brarian and
d dragged m
's where the
case you wo

```
2	2	CREEK .		1 2	0.5
10	7	I glanced and saw as bookcase went .	1 2 4 5 6 8 9	0.5
2	2	CRASH .		1 2	0.5
4	2	They stopped .	1 3	0.5
8	4	I bent my head .	1 2 3 4	0.5
9	5	What I do it cool .	1 3 4 7 8 0.5
14	9	I strutted and shut her mouth with my hand . 1 2 7 8 9 10 11 12 13	0.5
9	4	I slammed the book .	1 2 3 4	0.5
8	4	I check book .	3 4 7 8	0.5
12	8	" She glared at me and dragged me .	1 2 3 4 5 6 7 8	0.5
10	5	She backed me up .	1 2 3 4 10	0.5
7	6	`` Sit down and explain yourself .	1 2 3 4 5 6	0.5
5	2	He looked .	2 3	0.5
7	4	That 's story ends .	1 2 5 6	0.5
11	9	What I was about to tell him you heard .	1 2 3 4 5 6 7 8 10 0.5
16	12	I do n't want to be skeleton by the time I blabbing .	1 2 3 4 5 6 8 9 10 11 12 15	0.5
14	9	You want to know I will be mopping floors .	4 5 6 7 8 9 10 11 13	0.5
16	8	I got to go my friend calling me .	3 5 6 7 9 11 14 15	0.5
```

Kate skimme ... he li
brarian took ... fore .
They be books I thought .I went to have a look at them .One of the ones glittered .I tried to take it to look at it .Book seemed to be stuck .I stomped my foot .I picked my gum of my mouth and stuck it .Tried again tug book but it n't want to leave .I felt something gushy .I was losing my patience .I had time .I gritted my teeth and lifted my foot .Can you believe it it was my GUM .I put my shoe on the ground .They send me bill I have time .I was stubborn , as I mentioned before .I tugged at it time .I had made my decision .I was leaving with book even if I had to take shelf .CREEK .I glanced and saw as bookcase went .CRASH .They stopped .I bent my head .What I do it cool .I strutted and shut her mouth with my hand .I slammed the book .I check book ." She glared at me and dragged me .She backed me up .`` Sit down and explain yourself .He looked .That 's story ends .What I was about to tell him you heard .I do n't want to be skeleton by the time I blabbing .You want to know I will be mopping floors .I got to go my friend calling me .

FIGURE 4.7: Sentence Scoring Window

The following is the example to compare original sentence with the compressed sentence by human and Sentence Compressor:

Original sentence: The girl saw a cat at the backyard.

Gold standard: The girl saw a cat.

Human compression: the girl saw a cat.

Sentence Compressor output: The girl saw cat.

4.3 Experiment

The evaluation for Sentence Compressor is focusing on the summarized output produce by this application. The similarity of the summarized output is measured with the gold standard for every article. Five articles are selected from Astro Awani English news website. The selected articles were given to the lecturers from humanities department who have English certificate and qualification. They have to shorten every long sentence in the article. They do not have to shorten the sentence that should remain as it is (the sentence already in shortest form or the original meaning will be changed if the sentence is shortened). The shortened article written by them is set as the gold standard for evaluation.

After that, another five random respondents are selected. They were given the same previous task. They have to reduce the length of the long sentences from the article. The shortened article is recorded. The compressed output of the article is then produced by the Sentence Compressor. The articles were compressed separately. The five articles are included in Appendix section.

In this experiment, Bilingual Evaluation Understudy (BLEU) is used to measure the similarity of candidate and reference texts. First, the similarity of the articles compressed by human is measured with the gold standard. After that, the similarity of the articles compressed by the Sentence Compressor is measured with the gold standard. Figure 4.8 shows the BLEU scores for the five articles produced by human and Sentence Compressor.

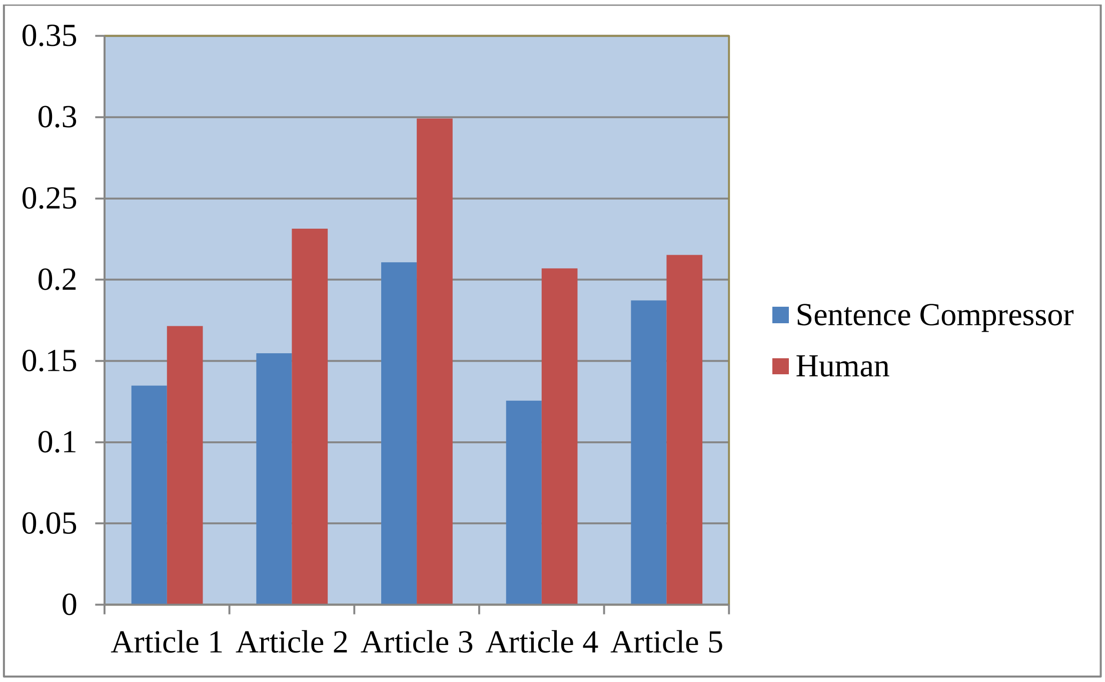


FIGURE 4.8: BLEU Scores

The BLEU score should be between 0 and 1. The highest the BLEU score, the higher the similarity of the text with the referenced text (gold standard). From the results produce, the compressed articles produced by Sentence Compressor are less similar to the gold standard compare to the compressed article by human. It can be assumed that the output by Sentence Compressor is less accurate compare to the output by human compression. A lot of improvements need to be done in order to reduce the score differences between these two types of output.

4.4 System Performance Evaluation

User testing had been conducted to evaluate the system's performance. 20 students from Universiti Teknologi PETRONAS had been randomly selected to test this Sentence Compressor. The testing was divided into two sessions.

In the first session, every respondent was given an article with long sentences before they try the Sentence Compressor. They were asked to read the whole article. The time for the respondents to finish the reading was recorded. After that, they were asked to list down the important keywords within 30 seconds. The number of keywords extracted from article is recorded.

In the second session, they were given a new article that contains of long sentences, but has the same length as the previous article. Then, they have to compress the sentences from the article by using the Sentence Compressor. After that, the steps in the first session are repeated where the respondents have to read the compressed article and find the important keywords. The time taken to read the article and the number of keywords extracted within 30 seconds is recorded.

At the end of the evaluation session, the respondents were given a form to evaluate the Sentence Compressor Application. Figure 4.8 shows the background of study for every respondent according to gender.

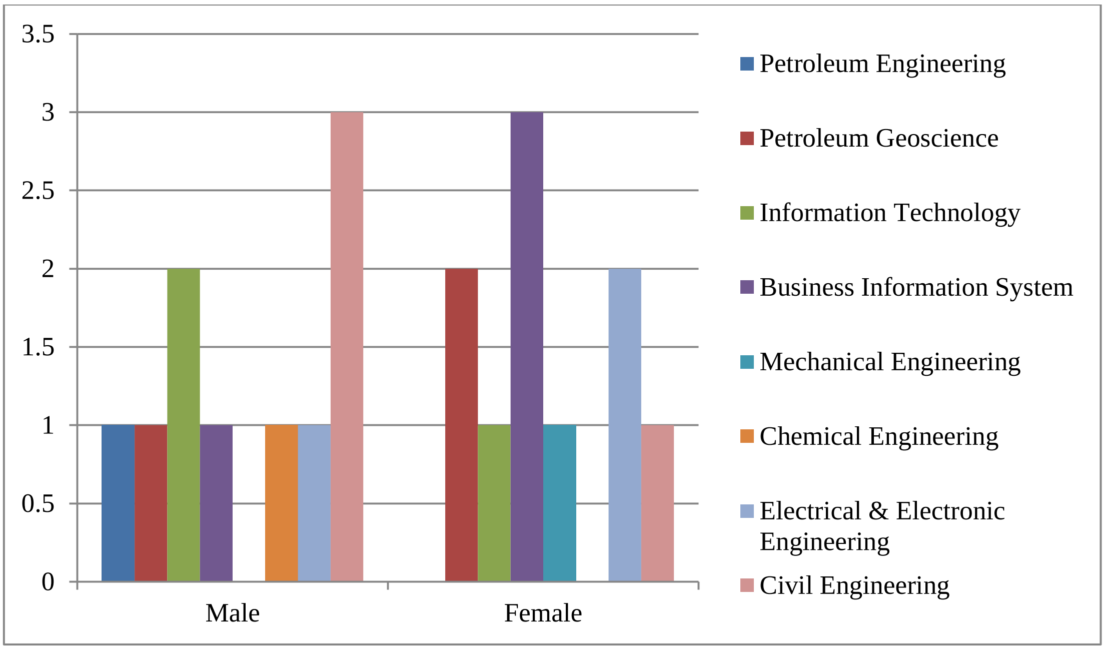


FIGURE 4.9: Respondents' Background of Study According to Gender

The respondents were given three statements to be evaluated with five options of answers which are strongly agree, agree, neutral, disagree, strongly disagree. The statements are:

- Sentence Compressor increases the readability of the text.
- It is easy to find keywords from compressed sentences.
- Sentence Compressor is very useful for information searching.

Figure 4.13 shows the results of the survey for 20 respondents. 70% of the respondents strongly agree and agree that the compressed sentence increase the readability of the text. Only two persons from the respondents strongly disagree that the Sentence Compressor increases the readability of the text. 65% of the users strongly agree and agree that it is easy to find the important keywords from the shortened sentence. 65% of the users strongly agree and agree that this application is helpful for information searching. 10% of the respondents strongly disagree that Sentence Compressor is very useful in information searching.

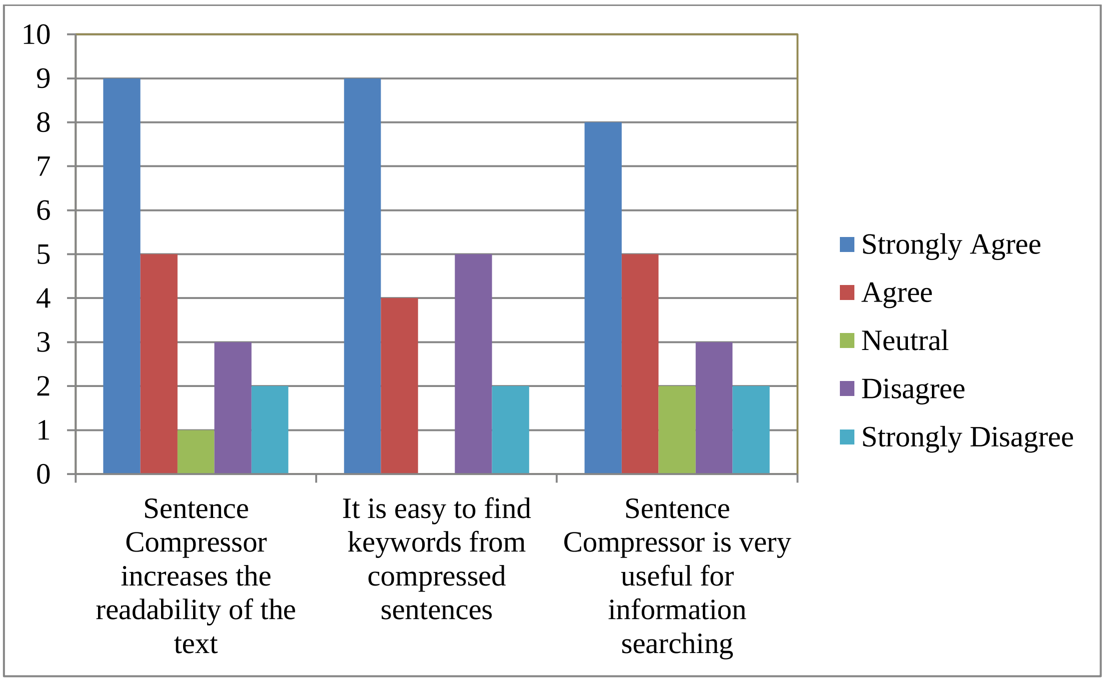


FIGURE 4.10: System Performance Evaluation Results

Based on this evaluation results, it is concluded that Sentence Compressor is useful in information searching since it can cut the time consumed to find the key information. Most of the users agreed that this application increase the readability of the text and make the keyword searching task easier.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

Sentence Compressor is an application that aims to reduce the length of the long sentences from text article. Most of the existing text summarizer available on the internet is focus on extraction-based summarization where it extracts the most important sentences in the document and combines them as the summary. This project aims to reduce the length of the long sentences since some of the sentences contain unimportant words.

The study proved that the summarized article and content is more readable and increase the efficiency of mobile learning. The summarized article will reduce the effort and time of the mobile user to find the important article.

This project focuses on sentence compression which is a part of abstraction-based summarization. The procedures to reduce the length of the sentences are divided into four phases. First phase is keywords extraction. Second phase is calculating the objective function. The third phase is adding the constraints and last but not least compressing sentence process.

The sentence compression by using Integer Linear Programming (ILP) approach is implemented as a desktop application that reduces the length of the long sentences to shorter ones by calculating the objective function and considering the constraints.

As the recommendation, this project should be continued by adding some new rules such as substitution rule and paraphrasing rule.

REFERENCES

[1] C. Smith. (2010, June). *Internet Usage Statistics: How We Spend Our Time Online (INFOGRAPHIC)* [Online]. Available: http://www.huffingtonpost.com/2010/06/22/internet-usage-statistics_n_620946.html

[2] Global Mobile Statistics 2013 Part B: Mobile Web; Mobile Broadband Penetration, 3G/4G Subscribers and Networks [Online]. Available: http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats/b

[3] R. Hema, "Natural Language Processing," in Proceeding of the National Seminar on Artificial Intelligence, Khentawas, Farrukh Nagar, 2009, pp. 89-98.

[4] C. Smith and A. Jönsson. "Automatic Summarization as Means of Simplifying Texts. An Evaluation for Swedish," Technical report, Santa Ann IT Research Institute.

[5] G. Yang, D. Wan, K., N. Chen and E. Sutinen, (2012). "Personalized Text Content Summarizer for Mobile Learning: An Automatic Text Summarization System with Relevance Based Language Model" in Proceedings of the 2012 IEEE Fourth International Conference on Technology for Education, pp. 90-97.

[6] I. Mani. "Summarization Evaluation: An Overview," Technical report, The MTRE Corporation.

[7] V. Yatsko, et al, "Automatic Genre Recognition and Adaptive Text Summarization," in Automatic Documentation and Mathematical Linguistics, 2010, Volume 44, Number 3, pp. 111-120.

[8] S. Angeles. (2013, February). *4 Things to Know Before Building an Android App* [Online]. Available: http://www.ideatoappster.com/4-things-you-need-to-know-before-building-an-anroid-app/

[9] O. Moknov. (2011, March). *20 New Frameworks for Web and Mobile App Developers* [Online]. Available: http://speckyboy.com/2011/03/07/20-new-frameworks-for-web-and-mobile-app-developers/

[10] *About the Project* [Online]. Available: http://phonegap.com/about/

[11] M. A. Rodis, (2013, May). *Stremor Corp's Liquid Helium Search and Summarization APIs Automatically Create Short Versions of Long Content* [Online]. Available: http://www.prweb.com/releases/summary/API/prweb10739704.htm

[12] U. Hahn and I. Mani (2000, November). "The Challenges of Automatic Summarization," in Cover Feature, 2000, pp. 29-35.

[13] I F. Moawad, M. Aref, (2012). "Semantic Graph Reduction Approach for Abstractive Text Summarization" in Computer Engineering & Systems (ICCES), 2012 Seventh International Conference, pp. 132-138.

[14] M.S. Binwahlan, N. Salim and L. Suanmali, (2009). "Swarm Based Text Summarization" in 2009 International Association of Computer Science and Information Technology – Spring Conference, pp. 145-149.

[15] T. Chang and W. Hsiao, (2008). "A Hybrid Approach to Automatic Text Summarization" in Computer and Information Technology, 2008. CIT 2008. 8th IEEE International Conference, pp. 65-69.

[16] P. Zhang and C. Li, (2009). "Automatic Text Summarization Based On Sentences Clustering And Extraction" in Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference.

[17] C.L. Devasena and M. Hemalatha, (2012). "Automatic Text Categorization and Summarization using Rule Reduction" in International Conference on Advances in Engineering, Science And Management (ICAESM -2012), pp. 594-598.

[18] L. Suanmali, N. Salim and M.S. Binwahlan, (2011). "Fuzzy Genetic Semantic Based Text Summarization" in 2011 IEEE Ninth International Conference on Dependables, Autonomic and Secure Computing, pp. 1184-1190.

[19] M.S. Binwahlan, N. salim and L. Suanmali, "Fuzzy Swarm Based Text Summarization" in • *Journal of Computer Science*, vol. 5, pp. 338-346, 2009.

[20] L. Chengcheng (2010). “Automatic Text Summarization Based On Rhetorical Structure Theory” in 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), pp. 595-598.

[21] M. Bonzanini, M. Martinez-Alvarez and T. Roelleke, (2013). “Extractive Summarisation via Sentence Removal: Condensing Relevant Sentences into a Short Summary” in 36th International ACM SIGIR Conference on Research and Development in Information Retrieval.

[22] F. Boudin and E. Morin, (2013). “Keyphrase Extraction for N-best Reranking in Multi-Sentence Compression” in The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013), pp. 298-305.

[23] *Summary: Procedures* [Online]. Available: http://www.kfmaas.de/summary0.html

[24] *English Grammar: Sentence Construction* [Online]. Available: http://www.learnenglish.de/grammar/sentencetext.html

[25] S. Bird, E. Klein, and E. Loper (2009). *Natural Language Processing with Python: Analysing Text with Natural Language Toolkit* [Online]. Available: http://nltk.org/book/ch08.html

[26] J. Clarke and M. Lapata, (2008). “Global Inference for Sentence Compression An Integer Linear Programming Approach” in journal of Artificial Intelligence Research, vol. 31, pp. 339-429, 2008.

[27] W.L. Winston and M. Venkataramanan, (2003). *Introduction to Mathematical Programming: Applications and Algorithms* (4th edition). Duxbury

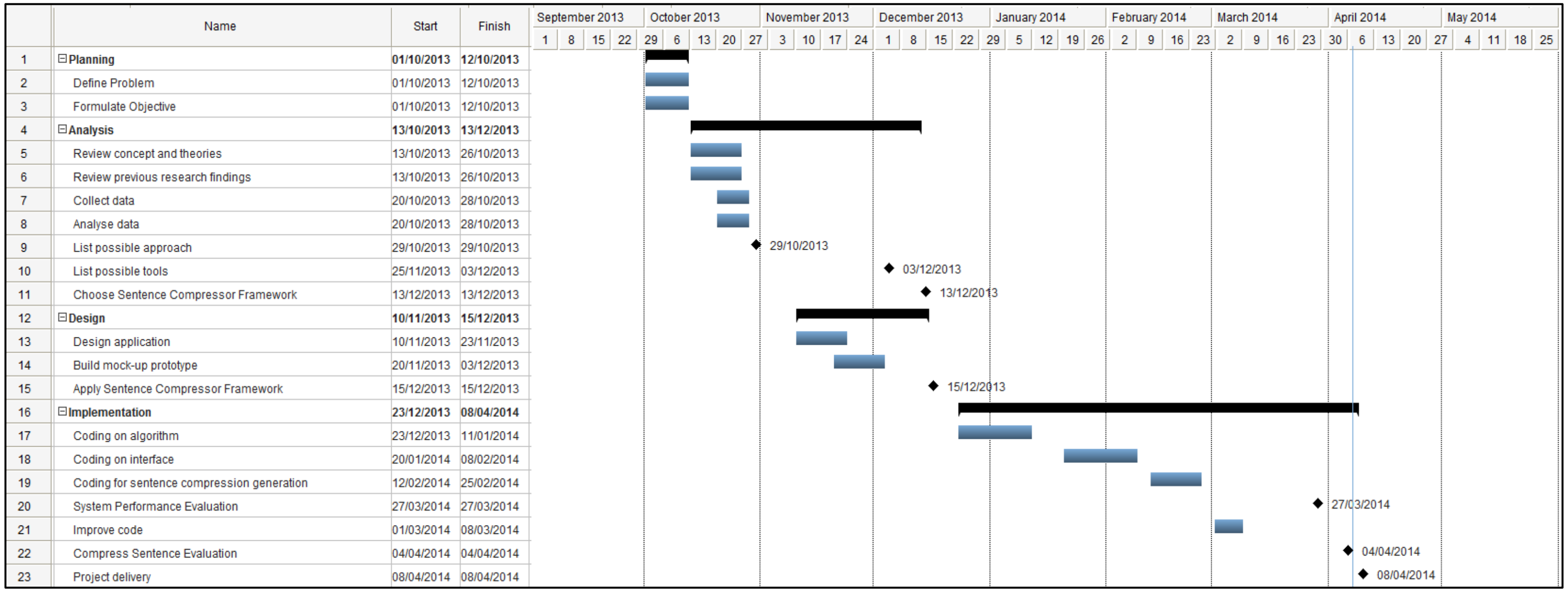
[28] K. Knight and D. Marcu, (2002). “Summarization beyond sentence extraction: a probabilistic approach to sentence compression” in journal of Artificial Intelligence, vol. 139, pp. 91-107.

[29] H. Jing, (2000). “Sentence reduction for automatic text summarization” in Proceedings of the 6th Applied Natural Language Processing Conference, Seattle, WA, USA, pp. 310-315.

[30] C. Hori and S. Furui, (2004). “Speech summarization: an approach through word extraction and a method for evaluation” in journal of IEICE Transaction on Information and Systems, vol. E87-D, pp.15-25.

[31] J. Turner and E. Charniak (2005). “Supervised and unsupervised learning for sentence compression” in Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, Ann Arbor, MI, USA, pp. 290-297.

APPENDIX I



	Name	Start	Finish
1	⊟**Planning**	**01/10/2013**	**12/10/2013**
2	Define Problem	01/10/2013	12/10/2013
3	Formulate Objective	01/10/2013	12/10/2013
4	⊟**Analysis**	**13/10/2013**	**13/12/2013**
5	Review concept and theories	13/10/2013	26/10/2013
6	Review previous research findings	13/10/2013	26/10/2013
7	Collect data	20/10/2013	28/10/2013
8	Analyse data	20/10/2013	28/10/2013
9	List possible approach	29/10/2013	29/10/2013
10	List possible tools	25/11/2013	03/12/2013
11	Choose Sentence Compressor Framework	13/12/2013	13/12/2013
12	⊟**Design**	**10/11/2013**	**15/12/2013**
13	Design application	10/11/2013	23/11/2013
14	Build mock-up prototype	20/11/2013	03/12/2013
15	Apply Sentence Compressor Framework	15/12/2013	15/12/2013
16	⊟**Implementation**	**23/12/2013**	**08/04/2014**
17	Coding on algorithm	23/12/2013	11/01/2014
18	Coding on interface	20/01/2014	08/02/2014
19	Coding for sentence compression generation	12/02/2014	25/02/2014
20	System Performance Evaluation	27/03/2014	27/03/2014
21	Improve code	01/03/2014	08/03/2014
22	Compress Sentence Evaluation	04/04/2014	04/04/2014
23	Project delivery	08/04/2014	08/04/2014

Gantt Chart

APPENDIX II

Article 1: US tiger kills Malayan female on first mate

LOS ANGELES: Zoo keepers are probing why a male Malayan tiger killed a four-year-old female tiger it had only just met, when they were brought together for breeding purposes in California.

Connor, who was born at San Diego Zoo in 2011, was introduced to female Tiga Tahun on Saturday for their first mating session, and initially everything seemed to be going routinely.

"Unfortunately this changed quickly and zookeepers were unable to separate the two animals," said zoo spokeswoman Jennifer Mehlow, explaining how the male tiger suddenly displayed "aggressive behavior."

The female feline died of injuries to her neck and breathing ability inflicted by Connor, she said. The death occurred before the world-famous zoo opened, so no visitors witnessed the incident.

It was the first fatal attack during a mating session at the zoo, although large carnivores often mix aggressive behavior with breeding interactions, sometimes leading to injury, the spokeswoman said.

The Malayan tiger, who was born at the Bronx Zoo in New York, was brought to San Diego earlier this year for breeding purposes, in an effort to propagate the endangered species, Mehlow said.

Tigers are normally solitary animals, and males and females only usually meet to mate when the female is in season, according to the Los Angeles Times, which said Connor is now the zoo's only Malayan tiger.

The sub-species is listed as endangered because of poaching and loss of habitat in Malaysia and parts of Thailand where they are found, the newspaper said.

APPENDIX III

Article 2: China's one-child policy change starts first quarter of 2014

BEIJING: China is expected to relax its decades-long one-child policy in the first quarter of 2014, said the National Health and Family Planning Commission.

The change may be implemented early next year after local legislatures passed the amended regulation, director of the family planning instruction department of the Commission, Yang Wenzhuang was quoted by Xinhua news agency as saying.

The move aims to raise fertility rates and ease the financial burden of China's rapidly ageing population.

Minister in charge of the National Health and Family Planning Commission, Li Bin when briefing lawmakers on Monday warned that if the current family planning policy persisted, the birth rate would continue to fall and lead to a sharp drop in the total population after reaching a peak, the state-run news agency reported.

Zhai Zhenwu, director of the School of Sociology and Population Studies at Renmin University of China, estimated the relaxation would lead to a mini baby boom lasting five or six years with an additional two million births a year on average.

The loosened policy would allow 15 million to 20 million Chinese couples eligible to have a second child, he said, adding a survey conducted last year showed that more than 60 per cent of newly eligible families were willing to have a second baby.

However, National People's Congress Standing Committee member Chi Wanchun said easing the policy was unlikely to abandon the family planning policy.

"Rather, it is also a measure for family planning," he was quoted as saying.

APPENDIX IV

Article 3: Hospital chef burnt to death in car-trailer collision

SUNGAI PETANI: A 35-year-old cook who worked at the Sultan Abdul Halim Hospital (HSAH) was burnt to death when his car burst into flames after colliding with a plywood-laden trailer.

The victim, Mohd Shamshuri Osman, 34, was said to be on his way home at about 8pm last night when the his Proton Saga BLM was believed to have skidded into the opposite direction at Jalan Sungai Petani-Bedong, near a petrol station.

The vehicle then collided with the lorry's petrol tank, and subsequently burst into flames.

According to a witness, Arvin Sonamnathan, 18, Mohd Shamshuri was seen trapped at the driver's seat.

“I heard a sound of someone calling for help but could not do anything because by then the fire had spread to the whole car,” said Arvin.

It was understood that Mohd Shamshuri was headed towards Bedong, Kedah while the lorry was heading towards Sungai Petani carrying goods from Gurun.

Meanwhile, Sungai Petani fire station chief Rosdi Ismai said that eight firemen were dispatched to the scene to put out the fire and remove the charred remain.

The accident is being investigated under Section 41(1) of the Road Transport Act 1987.

The remains of Mohd Shamshuri was sent to the HSAH for post-mortem.

APPENDIX V

Article 4: How sticking needles in your ear could help with weight loss

A small study out of Korea has found that ear acupuncture could be used to help people shed belly fat.

In the same way that pressure points in the feet are supposed to be linked to certain organs in the body, a new study claims that the stimulation of five acupuncture points in the ear could help patients with weight loss, particularly in battling the midriff bulge.

Published online in Acupuncture in Medicine, scientists out of Kyung Hee University in Seoul divided participants into three groups: 31 people were given acupuncture treatments to the outer ear in five different points. The needles were inserted 2 mm deep into the flesh and held in place with surgical tape for a week. The same treatment was switched to the other ear, and the process repeated over eight weeks.

In the second group of 30 participants, a needle was applied at one single hunger point. The third group was given a “sham” treatment which involved removing the needles immediately after insertion.

Throughout the experiment, it should be noted that participants were asked to follow a reduced calorie diet and to refrain from exercise. In total, 24 people dropped out before the eight-week term was up, 15 of whom were in the control group.

APPENDIX VI

Article 5: Bank Negara formulates new measurement for BLR next year

KUALA LUMPUR: Bank Negara Malaysia (BNM) is formulating a new interest rate framework for the measurement of the Base Lending Rate (BLR) next year to further enhance the country's financial system.

Governor Tan Sri Dr Zeti Akhtar Aziz said BNM would issue a consultative paper outlining the new reference rate framework to the industry in early January next year, as it needs to adjust and progress towards better serving the economy.

"Of late, there have been indications that the BLR as a reference rate has become less relevant.

"It has therefore become less meaningful as a basis for the pricing of loans. Retail lending rates on new loans being offered by the industry are at a substantial discount to the BLR," she told reporters after the launch of the Asian Banking School here Wednesday.

Zeti said the BLR is the key element in the financial intermediation process and needs to be improved in term of its efficiency, for the banking industry to continue to meet the ultimate objectives for its existence.

APPENDIX VII

Sentence Compressor Performance Evaluation Form

Sentence Compressor is an application that reduces the length of the long sentences by deleting the unimportant words in the sentence. The main objective of this evaluation is to measure the usefulness of Sentence Compressor in information searching.

Gender:

□ Male □ Female

Background of Study:

□ Information Technology
□ Business Information System
□ Petroleum Engineering
□ Petroleum Geoscience
□ Civil Engineering
□ Chemical Engineering
□ Mechanical Engineering
□ Electrical & Electronic Engineering

Do you agree with the following statements?

1. Sentence Compressor increases the readability of the text.

□ Strongly Agree □ Agree □ Neutral □ Disagree □ Strongly Disagree

2. It is easy to find keywords from compressed sentences.

□ Strongly Agree □ Agree □ Neutral □ Disagree □ Strongly Disagree

3. Sentence Compressor is very useful for information searching.

□ Strongly Agree □ Agree □ Neutral □ Disagree □ Strongly Disagree