RELEVANCE DETECTION AND SUMMARIZING STRATEGIES IDENTIFICATION ALGORITHM USING LINGUISTIC MEASURES

SEYED ASADOLLAH ABDIESFANDANI

THESIS SUBMITTED AS FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA

2016

Abstract

Summarization is a process to select important information from a source text. Summarizing strategies are the core of the cognitive processes involved in the summarization activity. Summarizing strategies include a set of conscious tasks that are used to determine important information and extract the main idea of a source text.

In this research project, we conducted a study on students' summaries. The findings of the study show that, there is a strong relationship between the summary writing proficiency of students and the summarizing strategies that they used. We then develop a new algorithm to address the summarizing strategies identification problem. The algorithm simulates two important tasks that are frequently used by the human experts to identify summarizing strategies used to produce the summary sentences: 1) sentences relevance identification; and 2) summarizing strategies identification.

The sentences relevance identification module uses a statistical based approach such as vector space model (VSM) to represent sentences and compute similarity between the source sentences and the summary sentences using the cosine similarity measure. It then integrates both the semantic and syntactic similarity measures using a linear equation to capture the meaning in comparison between two sentences. It aims to distinguish the meaning of two sentences, when two sentences have same surface or share the similar bag-of-words (BOW), while their meaning is different. The module also employed a word semantic similarity measuring method to overcome vocabulary mismatch problem in sentence comparison. The method bridges the lexical gaps for semantically similar contexts that are expressed in a different wording. In addition, the sentences relevance identification module requires some degree of linguistic pre-processing, including part of speech tagging (POS), word stemming and stop-words removal.

The summarizing strategies identification module relies on a set of heuristic rules, statistical and linguistic methods such as position-based method, title-based method, cue-phrase method and word-frequency method to identify the summarizing strategies employed by students.

To evaluate the algorithm, we conducted two experiments. In the first experiment, we examine the functionality of the system, whether the system is able to identify the summarizing strategies used by students in summary writing. The result for the first experiment shows that the system is able to identify some of summarizing strategies which are deletion, sentence combination, paraphrase and topic sentence selection. The system is also able to detect copy- verbatim strategy, the most commonly strategy used by students. Besides than these strategies, there are four methods used in topic sentence selection strategy which can also be identified by the system. They are 1) cue method; 2) title method; 3) keyword method; and 4) location method. In the second experiment, we want to measure the performance of the algorithm against human judgment to identify the summarizing strategies using the precision, recall, F-measure score and accuracy rate. The experimental results show that the proposed algorithm achieved acceptable results in comparison to human judgment. The algorithm achieved an average of 87% precision, 83% of recall, 85% of F-score and 82% of accuracy rate.

Abstrak

Rumusan adalah satu proses untuk memilih maklumat penting dari teks sumber. Strategi-strategi ringkasan adalah teras kepada proses kognitif yang terlibat dalam aktiviti rumusan. Strategi-strategi ringkasan termasuk satu set tugas sedar yang digunakan untuk menentukan maklumat yang penting dan mengeluarkan idea utama bagi teks sumber.

Dalam projek penyelidikan ini, kami telah menjalankan kajian terhadap ringkasan pelajar. Hasil kajian menunjukkan bahawa terdapat hubungan yang kuat antara kemahiran pelajar menulis ringkasan dan strategi-strategi ringkasan yang digunakan. Selain itu, kami membangunkan satu algoritma baru untuk menangani masalah pengenalpastian strategi-strategi ringkasan. Algoritma mensimulasikan dua tugas penting yang sering digunakan oleh pakar-pakar manusia untuk mengenalpasti strategi-stratgei ringkasan yang digunakan untuk menghasilkan ayat ringkasan:1) pengenalpastian perkaitan ayat-ayat dan 2) pengenalpastian strategi-strategi ringkasan.

Modul pengenalan ayat-ayat relevan menggunakan pendekatan yang berdasarkan statistik seperti model ruang vektor (VSM) untuk mewakili ayat-ayat dan mengira persamaan antara ayat dari sumber teks dan ayat ringkasan menggunakan ukuran persamaan kosinus. Kemudian, ia menggabungkan kedua-dua ukuran persamaan semantik dan sintaksis menggunakan persamaan linear untuk mendapatkan makna dalam perbandingan antara dua ayat. Ia bertujuan untuk membezakan makna bagi dua ayat, apabila kedua-dua ayat tersebut mempunyai permukaan yang sama atau berkongsi perkataan-perkataan yang sama (BOW), tetapi membawa maksud yang berbeza berbeza. Modul ini juga menggunakan kaedah pengukuran persamaan semantik perkataan untuk mengatasi masalah ketidaksesuaian perbendaharaan kata dalam perbandingan ayat. Kaedah ini merapatkan jurang leksikal bagi secara semantiknya

iv

konteks yang sama yang dinyatakan dalam kata-kata yang berbeza. Di samping itu, modul ini juga memerlukan beberapa bahasa pra-pemprosesan, termasuk tag bahagian ucapan (POS), dasar perkataan dan penyingkiran *stop-words*.

Modul pengenalan strategi menulis ringkasan bergantung kepada satu set peraturan heuristik, kaedah statistik dan bahasa seperti kaedah frekuensi perkataan, kaedah berdasarkan kedudukan, kaedah berdasarkan tajuk dan kaedah petunjuk-frasa untuk mengenalpasti strategi membuat ringkasan yang digunakan oleh pelajar-pelajar.

Untuk menilai algoritma tersebut, kami telah menjalankan dua eksperimen. Dalam eksperimen pertama, kami memeriksa fungsi sistem, sama ada sistem mampu untuk mengenalpasti strategi-strategi ringkasan yang digunakan oleh pelajar dalam penulisan ringkasan. Keputusan eksperimen pertama menunjukkan bahawa sistem ini mampu mengenalpasti beberapa trategi rimgkasan seperti penghapusan, kombinasi ayat, parafrasa dan pemilihan ayat topik. Sistem ini juga mengenalpasti strategi salin kata demi kata. Selain daripada strategi yang dinyatakan, terdapat empat kaedah yang digunakan bag pemilihan ayat topic iaitu: kaedah isyarat, kaedah tajuk, kaedah kata kunci dan kaedah lokasi, yang juga boleh dikenal pasti oleh sistem. Dalam eksperimen kedua, kami mahu mengukur prestasi algoritma terhadap penghakiman manusia untuk mengenal pasti strategi ringkasan menggunakan *Precision*, *Recall*, *F-measure score* dan kadar ketepatan. Keputusan eksperimen menunjukkan bahawa algoritma yang dicadangkan mencapai keputusan yang boleh diterima dalam perbandingan penghakiman manusia. Algoritma tersebut mencapai purata 87% *Precision*, 83% *Recall*, 85% *F-measure score* dan 82% kadar ketepatan.

Acknowledgements

First and foremost, I thank Allah for guiding me and taking care of me all the time. My life is so blessed because of his majesty.

I wish to express my deep appreciation to my research supervisor, *Dr. Norisma Benti Idris*, for all her guidance, continuous support and enthusiasm throughout this research and other courses. Her valuable guidance aided me immensely during the preparation of this thesis. No matter how busy she has been, she has always found time to answer my frequent questions and solve the problems that I came across during my PhD program.

I would like to thank Professor Dr. Sapiyan Baba, Dr. Rukaini Haji Abdullah, Dr. Rohana Mahmud and Dr. Ram Gopal Raj for providing insightful comments on my thesis and presentations. I would also like to take this opportunity to thank all my friends who, in one way or another, have helped with the process of doing this research.

I would also like to give my special thanks to Professor Dr. Ramiz M. Aliguliyev and Professor Dr. Rasim M. Alguliyev from the Azerbaijan National Academy of Sciences in Azerbaijan for the interesting discussions and collaborations that we had. Their directions gave me valuable knowledge on text summarization.

Finally, i wish to express my considerable gratitude to my parents and family for their moral support and inspiration which have brought me here today.

March, 2016

vii

Table of Contents

CHAPTER 5: RELEVANCE DETECTION AND SUMMARIZING

List of Figures

List of Tables

List of Abbreviations

AR	:	Array Root
AS	:	Array Synonym
BOW	:	Bag of Words
BLEU	:	Bilingual Language Evaluation Understudy
BP	:	Brevity Penalty
CAA	:	Computer Assisted Assessment
CWL	:	Cue Word List
DM	:	Discourse Marker
IC	:	Information Content
IDF	:	Inverse Document Frequency
KL	:	Keywords List
LCS	:	Longest Common Subsequence
LSA	:	Latent Semantic Analysis
NLP	:	Natural Language Processing
NSS	:	Number of Summarizing Strategies
NFQA	:	Non Factoid Question Answering
POS	:	Part-of-Speech
ROUGE	:	Recall Oriented Understudy for Gisting Evaluation
RW	:	Root of Word
RDSSIA	:	Relevance Detection and Summarizing Strategies Identification Algorithm
SRDS	:	Sentences Relevance Detection Stage
SS	:	Summary Sentence
SSBW	:	Semantic Similarity Between Words
SSCS	:	Sentence Similarity Computation Stage

SSCM	:	Sentence Similarity Computation Model
SSDS	:	Summarizing Strategies Detection Stage
SSBS	:	Semantic Similarity Between Sentences
SSM		Sentence Similarity Measurement
SVD	:	Singular Value Decomposition
SLL	:	Sentence Location List
SP	:	Students' Performance
SC	:	Sentence Combination
TF	:	Term Frequency
TFIDF	:	Term Frequency , Inverse Document Frequency
TL	:	Title List
TRDS	:	Text Relevance Detection Stage
TTS	:	Topic Sentence Selection
VSM	:	Vector Space Model
WS	:	Word Set
WOSBS	:	Word Order Similarity Between Sentences

List of Appendices

CHAPTER 1

INTRODUCTION OF THE STUDY

1.1 Introduction

Reading skills are essential for success in society. Reading affects different aspects in our life, especially in school. The aim of reading is to elicit meaning from the written text; hence, lack of capacity in this area may affect comprehension ability. Comprehension contains inferential and evaluative thinking, not just a reproduction of the author's words. In school, students' comprehension skills can be taught and improved during their learning process.

There are various forms of teacher-student discussions to improve comprehension ability (Barry, 2002; Fialding & Pearson, 1994), including where the teacher initiates a question, a student responds, and the teacher evaluates the response such as a multiple-choice question, true-false question and short answer question. According to the results of some researches, summarization can also be one of the main keys to improve reading comprehension. The purpose of summarization is to improve reading comprehension (Duke & Pearson, 2008; Graham & Hebert, 2010; Karbalaei & Rajyashree, 2010; Kashef, Damavand, & Viyani, 2012; Selinger, 1993).

Summarization is a process of automatically producing a compressed version of a given text that provides useful information for the user (Aliguliyev, 2009; Chatterjee & Sahoo, 2015; Galgani, Compton, & Hoffmann, 2014; John & Wilscy, 2015; Steinberger, Poesio, Kabadjov, & Ježek, 2007; Yang, Wen, & Sutinen, 2013). In addition, it is a process that involves several activities such as comprehension, selection, interpretation, transformation and generation. The main goal of summary writing

operation is to create a summary text. Summarizing instructs students concerning how to recognize the main ideas in a text, determine important information that is worth noting and eliminate irrelevant information (Brown & Day, 1983; Chang, Sung, & Chen, 2002; Wormeli, 2005; Zipitria, Arruarte, & Elorriaga, 2010; Zipitria, Elorriaga, Arruarte, & de Ilarraza, 2004). Summarization is a cognitive process to condense a text into its most important concepts, while, summarizing strategies are the core of the cognitive processes involved in the summarization activity (Kintsch & Van Dijk, 1978; Pakzadian & Rasekh, 2013). Summarizing strategies include a set of conscious tasks that are used to create a summary text. There are several summarizing strategies for determining and eliminating irrelevant information, and extracting the main idea of a source text. According to the result of some studies, a major difficulty faced by students in summary writing is the lack of skills in applying summarizing strategies (Huang, 2006; Idris, Baba, & Abdullah, 2009; Karbalaei & Rajyashree, 2010; Winograd, 1984; Zafarani & Kabgani, 2014). Since summarization is an important tool for improving comprehension and can be used as a measure of understanding in school (Chiu, Wu, & Cheng, 2013; Pressley, 1998; Westby, Culatta, Lawrence, & Hall-Kenyon, 2010), it has garnered a lot of interest from the teachers to teach summary writing through direct instruction (Casazza, 1993; Cho, 2012; Guido & Colwell, 1987; Hare & Borchardt, 1984; Hill, 1991; Taylor, 1986; Westby et al., 2010).

In direct instruction, teachers need to possess some information such as what summarizing strategies used by students, the ability of students to use summarizing strategies, and the students' weakness in summarizing. To collect all the information manually is difficult as it is a highly time consuming task. Hence, as one of the ways to reduce the time they should spend on this task, many teachers choose to reduce the number of summaries given to their students. This would cause students to have insufficient practice on summary writing, which undeniably affects their summary

2

writing ability (Y. He, Hui, & Quan, 2009). To tackle these problems, computer-assisted assessment (CAA), which has garnered much interest in recent years, is one of the methods that can be used to assist teachers. Due to the progress in other areas, such as E-learning, Information Extraction and Natural Language Processing, the automatic evaluation of summary writings has been made possible. Although previous systems have been developed to assess summary writings, most of them focus only on content coverage. Only a few systems have been developed to identify summarizing strategies used by students.

This research aims to develop an algorithm for the summarization assessment system that can be used to – first, detect text relevancy of students' summaries and secondly, identify the summarizing strategies employed by students in summary writing. Finally, it aims to provide teachers and students with a learning environment that can help them to identify summarizing strategies, produce their summaries with more quality and improve their comprehension.

It is worth noting that this work is not concerned with the summarization process, for which the result is a summary text, but with the summarization assessment process, for which the result is identifying summarizing strategies and detecting text relevancy of students' summaries.

1.2 Research Motivation

We focus on summarizing strategies because of several reasons. These reasons are as follows:

- **The educational benefits of summarization:**

Summarization training improves the quality of students' summaries (Brown, Campione, & Day, 1981; Cunningham, 1982; Hare & Borchardt, 1984) and it also has effects on reading comprehension measures (Baumann, 1984; Bean & Steenwyk, 1984; Chiu et al., 2013; Kashef et al., 2012; McNeil & Donant, 1982; Rinehart, Stahl, & Erickson, 1986). Often, direct instruction has been linked with teaching students on how to use a set of summarizing strategies or cognitive rules for summarizing. The direct instruction helps students to learn how to determine the main ideas of a source text, it also enables students to focus on key words and phrases of the assigned text that are worth noting and it teaches students how to reduce the text to its main points. The findings from these studies have attracted interest from the teachers for training summarizing strategies through instruction. To do so, they need to review and assess the students' summaries. If they want to do it manually, it can be overwhelming. This is where a computer-based system such as our proposed algorithm would be an advantage for the teachers.

The proposed algorithm is called RDSSIA: **R**elevance **D**etection and **S**ummarizing **S**trategies **I**dentification **A**lgorithm.

- **To develop a system into automated summarization assessment**

Most of the existing systems focused only on the quality of the summary, which are: content and style. Only a few systems focused on how to identify summarizing strategies.

- **To give an informative feedback to teachers and students**

> Identifying the strategies used by students in summary writing and knowing how much the information in the summary text overlaps with information in the source text can help both teachers and students. For the teacher it provides evidence of the student's ability to select the important information of a text. It provides evidence of the student's ability on how to use summarizing strategies. For the students, it provides a supportive learning environment which will help them improve their summarizing skills. The students can be taught to use the appropriate strategies for creating a good summary.

1.3 Problem Statement

Conceptually, the process of identifying summarizing strategies involves two sub-processes as shown in Figure 1.1. The processes are: 1) identifying the sentences from the source text that were used to create the summary sentences; and 2) identifying the summarizing strategies based on the sentences that have been identified in the first process. Before identifying the summarizing strategies, the Text Relevance Detection Stage (TRDS) should be able to determine the relevant sentences from the source text, for each summary sentence. If the relevant sentences cannot be determined from the source text, no matter how well other stages in the system perform, the summarizing strategies will not be identified.

Therefore, the text relevance detection Stage is an important engine in identifying summarizing strategies. This module provides a list of sentences which will be analysed in further steps. These sentences are then further processed using a variety of techniques to identify the summarizing strategies that has been used in summary writing.
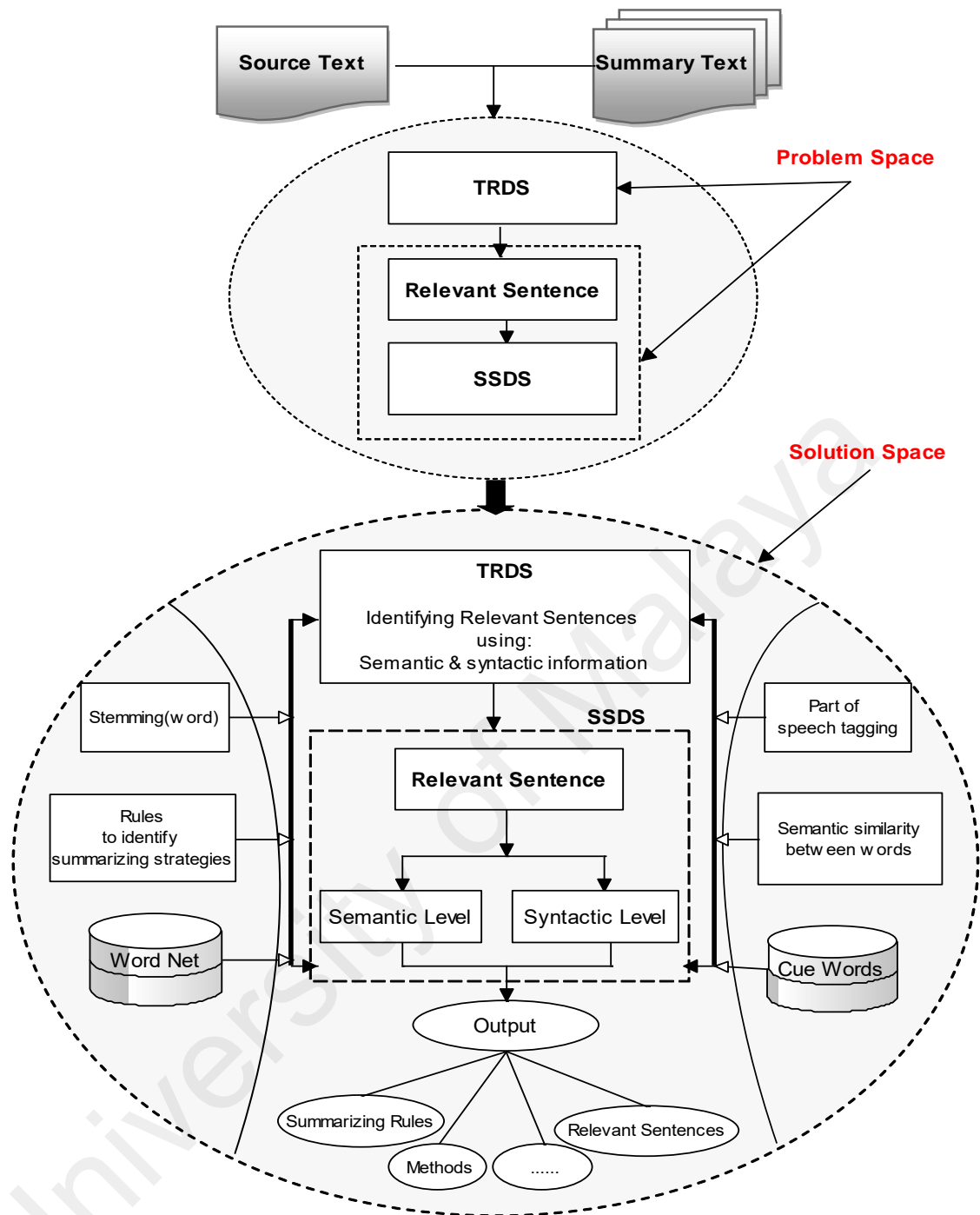
TRDS: Text Relevance Detection Stage

SSDS: Summarizing Strategies Detection Stage

Methods: (Cue,Title,Location,Key word)

Summarizing Rules: (Paraphrase,Deletion,Sentence Combination, Topic Sentence Selection,Copy

Word Net : Word Net is a lexical database for English which was developed at Princeton University

Cue Words : a list of discourse markers

Figure 1.1: Problem Space and Solution Space

In the context of text relevance, linguistic knowledge such as semantic relations between words and their syntactic composition, play key role in sentence understanding. This is particularly important in comparison between two sentences where a single word token is used as a basic lexical unit for comparison.

Syntactic information, such as word order, can provide useful information to distinguish the meaning of two sentences, when two sentences share the similar bag-of-words. For example, "*student helps teacher*" and "*teacher helps student*" will be judged as identical sentences because they have the same surface text. However, these sentences convey different meanings. On the other hand, two sentences are considered to be similar if most of the words are the same or synonyms. However, it is not always the case that sentences with similar meaning necessarily share many similar words. Hence, semantic information such as semantic similarity between words and synonym words can provide useful information when two sentences have similar meaning, but they used different words in the sentences.

On the other hand, while both semantic information and syntactic information contribute in sentence understanding (Achananuparp, Hu, & Shen, 2008; He, Li, Shao, Chen, & Ma, 2008; Kanejiya, Kumar, & Prasad, 2003; Pérez et al., 2005; Wiemer-Hastings & Wiemer, 2000; Wiemer-Hastings & Zipitria, 2001; Zhao & Tang, 2010), the current systems that have been proposed to identify summarizing strategies, did not use the combination of semantic relations between words and their syntactic composition to identify text relevancy. Obviously, this drawback has a negative influence on the performance of the previous systems.

As shown in Figure 1.1, there are two levels of summarizing strategies – semantic and syntactic levels. The strategies in semantic level include paraphrasing, generalization, topic sentence selection and invention. The strategies in syntactic level include deletion,

copy verbatim and sentence combination. A few systems have been proposed to identify summarizing strategies(Idris et al., 2009; Lemaire, Mandin, Dessus, & Denhière, 2005). However, these systems can either identify summarizing strategies at semantic level or syntactic level.

1.4 Aim and Objectives

The main goal of this research is to develop an algorithm that can be used to detect text relevancy of students' summaries and to identify the summarizing strategies employed by the students. To achieve this main goal, the following specific objectives are defined:

i. To compare the students' performance in summary writing with the summarizing strategies that they used.

ii. To develop an algorithm that can detect text relevancy and identify students' summarizing strategies.

iii. To compare the performance of the proposed algorithm with human judgement in order to increase the ratio of precision, recall and F-measure measurements for identifying summarizing strategies.

1.5 Research Questions

In principle, this thesis attempts to answer several research questions corresponding to the objectives identified in the previous section (refer to section 1.4).

i. **Objective 1:** To compare the students' performance in summary writing with the summarizing strategies that they used.

a) Is there a correlation between summarizing strategies and students' performance in summary writing?

b) Does the number of summarizing strategies that the students used, affect the students' performance in summary writing?

ii. **Objective 2:** To formulate an algorithm that can detect text relevancy and identify students' summarizing strategies.

    a) How can the relevancy between summary sentences and the sentences from the source text be detected?

    b) How can the summarizing strategies be identified?

    c) How can algorithm to detect text relevancy and identify summarizing strategies of students' summaries be formulated?

    d) How does the algorithm work?

iii. **Objective 3:** To compare the performance of the proposed algorithm with the human judgement in order to increase the ratio of precision, recall and F-measure measurements for identifying summarizing strategies.

    a) Can the proposed algorithm identify the summarizing strategies used by students?

    b) How is the performance of the algorithm when compared to human judgment?

1.6 Research Methodology

The research process involved the following five phases:

1) Problem analysis and collecting wide knowledge of summarization

Several areas related to the research objective were reviewed with regards to their possible contribution to RDSSIA process development. In total, the following fields of research were investigated:

- Automatic text Summarization

- Approaches to text summarization

- Summarization assessment

- Summary assessment techniques

- Various tools in summarization assessment

- Macro rules in summarization

- General rules for producing a summary

- Sentence Similarity Measures

Based on the analysis of these areas, the Algorithm, RDSSIA, for text relevance detection and summarizing strategies identification was developed. The problem space, solution space and the links between them are illustrated in Figure 1.1. They are derived using the following steps:

- The problem space focused on two main problems. First is the TRDS, where the relevant sentences are identified based on either semantic or syntactic similarity, not both. On the other hand, the second problem is the SSDS, where the summarizing strategies are identified in either semantic level or syntactic level.

- The solution space, the TRDS identified relevant sentence based on a combination of semantic relations between words and syntactic composition. The SSDS identifies summarizing strategies in both semantic and syntactic levels.

2) Collecting data and Data analysis

To analyse students' summarizing strategies samples of student-written summaries were collected. The samples will then be analysed to provide answers for these questions:

- The summarizing strategies used by students for producing a summary text.
- The correlation between summarizing strategies and students' performance.
- Whether the number of summarizing strategies that the students used affects their performance.

Details of the analysis are discussed in chapter 3.

3) Heuristic rules for identifying summarizing strategies

In the current phase, a study has been done on human– written summary to collect a set of rules for identifying summarizing strategies that are used in producing a summary. Details of the study are presented in chapter 4.

4) Development of the algorithm

In this project, we propose an algorithm to identify text relevancy and summarizing strategies, which it is called RDSSIA. We formulate a set of rules into the RDSSIA to identify summarizing strategies in semantic and syntactic levels. We also identify the approach to determine relevancy between source sentences and summary sentences, in which this approach compares two sentences based on semantic relations between words and syntactic composition. The RDSSIA was implemented to show how our

11

proposed algorithm could be used to identify summarizing strategies and text relevancy (details in Chapter 5).

5) Evaluation of algorithm

This phase contains the experiments carried out and the results obtained by the proposed algorithm. Firstly, we carried out an experiment to determine whether the algorithm is able to identify the summarizing strategies. Next, we also conducted some experiments to evaluate the performance of the algorithm by comparing the results obtained by the algorithm with the human judgment.

1.7 Thesis Overview

The overall structure of the thesis is illustrated in Figure 1.2. The first three chapters present the background information on the domains that are related to this research. The subsequent four chapters in the thesis describe the research contribution of this thesis.
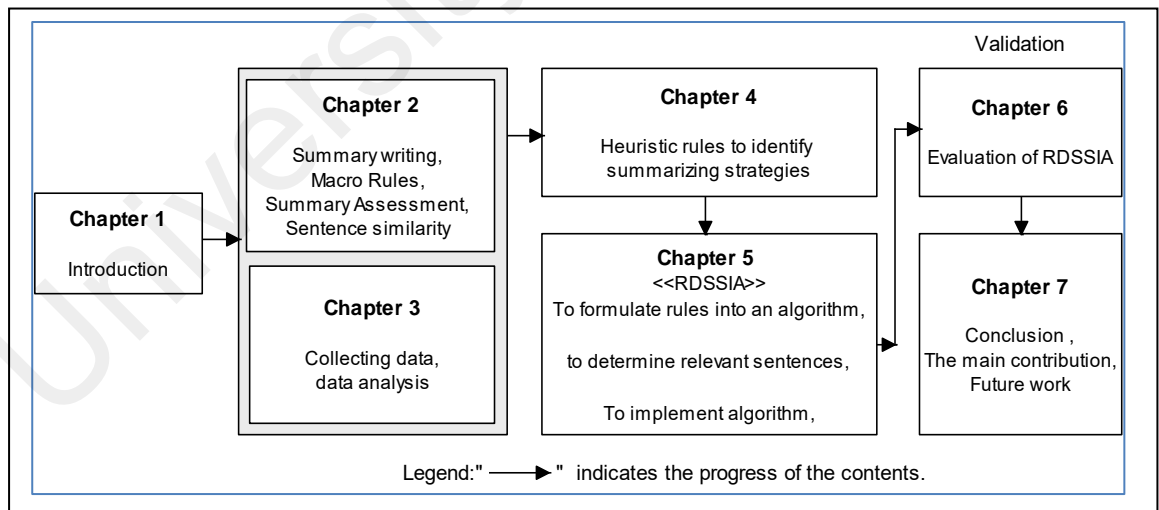
Figure 1.2: Overview of the Thesis

This thesis is organized as follows.

- Chapter 1 introduces the research topic and gives an overview of the research objectives, research questions, research motivation, research problem statement, research contribution and research methodology. It also presents the structure of the thesis.

- Chapter 2 gives a basic introduction into summarization. This chapter also clarifies the terminology used in summarization research and provides a description of summary evaluation metrics, macro rules and sentence similarity measure. It describes the problems that the current techniques encounter in identifying text relevancy and summarizing strategies and some methods that seem to be useful in tackling these problems.

- Chapter 3 describes specific considerations when dealing with summarizing strategies identification. This chapter specifically discusses the importance of summarizing strategies in students' performance.

- Chapter 4 presents an analysis on human-written summaries to determine a set of rules to identify summarizing strategies automatically. In this chapter several rules are explored to identify each summarizing strategies. The main contribution of this chapter is to provide the answer to the question; "*How can the summarizing strategies be identified?*"

- Chapter 5 describes the heart of the RDSSIA; it shows how semantic relations between words and syntactic composition can be utilized in text relevancy detection and summarizing strategies identification. This chapter also demonstrates how RDSSIA is able to identify relevant sentences and summarizing rules in semantic and syntactic levels.

- Chapter 6 includes evaluation results of the RDSSIA. This chapter presents evaluation results of two experiments. The first experiment is to show the functionality of the algorithm, RDSSIA, in identifying the four summarizing strategies such as deletion, sentence combination, paraphrase and topic sentence selection, and four methods: cue method, title method, keyword method and location method. The algorithm also identified copy-verbatim strategy, although this strategy is not part of summarizing strategies but it is used by students. The second experiment is to evaluate the performance of the algorithm when compared to human judgment.

- Chapter 7 presents the main conclusion of this research work and the main contribution of this thesis. It also addresses some issues that must be taken into consideration by future studies.

CHAPTER 2

TEXT SUMMARIZATION

2.1 Introduction to Text Summarization

It is generally agreed that well-developed reading comprehension ability is the key to students' academic success. This comprehension ability is not a passive state which one possesses, but it is an active mental process which needs to be improved. Students' comprehension skills can be improved during their learning process.

In traditional teacher-student discussions, the teacher initiates a question, a student responds, and the teacher evaluates the response. Recent studies show that various forms of teacher-student discussions try to achieve the following three goals (Barry, 2002; Pearson & Fielding, 1991):

- Embedding strategy instruction in text reading.
- Accepting personal interpretations and reactions.
- Changing teacher-student interaction patterns.

According to previous literatures, summarization is one of the important keys in reading comprehension and teaching. The purpose of summarization is to improve reading comprehension (Kashef et al., 2012; Selinger, 1993). Summarization is also a technique to improve students' reading comprehension skills (Alyousef, 2006; Brown & Day, 1983; Cho, 2012; Fan, 2010a; Hedge, 2001; Kamhi-Stein, 1993; Pakzadian & Rasekh, 2013; Zipitria et al., 2010; Zipitria et al., 2004).

The effects of summarization instructions on text comprehension – summarization can be used in teaching (Bartlett & Burt, 1933; Garner, 1982; Kintsch, Patel, & Ericsson, 1999; Zipitria et al., 2010) as an educational strategy to derive comprehension. "Practice in summarizing improves students' reading comprehension of fiction and nonfiction alike, helping them constructs an overall understanding of a text, story, chapter, or article" (Rinehart et al., 1986).

Previous studies have shown that summarization is the most effective teaching strategy in the history of education (Marzano, 2003, 2006; Marzano, Frontier, & Livingston, 2011). The aim of summarization instruction is to focus on the main idea, key details, key-words, phrases and to write adequately and simply but take complete notes. It also reduces the reading time (Mani et al., 2002).

2.2 Summarization

The main idea of summarizing process is to reduce the size and content of the source text into important information. The process contains the combination of information and the designation of the grade of importance of the information included in a text. In addition, it is a process that merges several activities such as comprehension, selection, interpretation, transformation, and generation. The main goal of summary writing operation is to create a summary. Unlike other types of writing such as report writing, the construction of summary depends on existing text and the summarizer's intention on what to comprise, what to delete, how to arrange information and how to certify that the summary is not changing the meaning of the original text (Cai, Li, & Zhang, 2014; Chen & Chen, 2012; Glavaš & Šnajder, 2014; Gupta & Lehal, 2010a; Kazantseva & Szpakowicz, 2010; Sobh, Darwish, & Fayek, 2006; Yang, Chen, Sutinen, Anderson, & Wen, 2013).

2.2.1. Rules in Summarization

Macro rules include a set of conscious tasks that are used to create a summary text. There are several summarizing strategies employed to determine important information, eliminating irrelevant information, and extracting the main idea of a source text (Cho, 2012; Idris et al., 2009; Kintsch & Van Dijk, 1978; Pakzadian & Rasekh, 2013).

Different terminology was used to explain the summarizing strategies. Several summarizing strategies proposed by these authors (Brown & Day, 1983; Idris et al., 2009; Johnson, 1983; Kintsch & Van Dijk, 1978; Lemaire et al., 2005; Westby et al., 2010) to produce an appropriate summaries. We describe these strategies as follows:

- **Deletion**

To produce a summary sentence, deletion strategy is used to remove unnecessary information in the sentence of the source text. Unnecessary information includes trivial details about the topics such as examples and scenarios or redundant information containing the rewording of some of the important information.

- **Sentence Combination**

Sentence combination strategy is employed to merge two or more phrases from the source sentences. These sentences are usually merged using conjunction words, such as "*for", "but", "and", "after", "since", and "before"*.

- **Generalization**

The generalization rule is an act of replacing a general term for a list. There are two replacements. One is the replacement of a general word for a list of similar items, e.g. "*pineapple, banana, star fruit and pear"* which can be replaced by "*fruits"*.

The other one is the replacement of a general word for a list of similar actions, e.g. the sentences: *"Yang eats a pear,* and *Chen eats a banana",* can be replaced by: *"The boys eat fruits".*

- **Paraphrasing**

In the paraphrasing process, a word in the source sentence is replaced with synonym word (different words with the same meaning) in the summary sentence.

- **Topic sentence selection**

To produce a summary sentence, topic sentence selection strategy is used to extract an important sentence from the original text to represent the main idea of a paragraph. There are four methods to identify the important sentence:

i. **Key method**

The most frequent words in a text are the most representative of its content, thus a segment of text containing them is more relevant (Laura Alonso et al., 2004). Word frequency is a method used to identify keywords that are non-stop-words, which occur frequently in a document (Teufel & Moens, 1997; Xie & Liu, 2008, 2010). According to Gupta and Lehal (2010a), sentences with keywords or content words have a greater chance of being included in the summary.

ii. **Location method**

Important sentences are normally located at the beginning and the end of a document or paragraphs, as well as immediately below section headings (Fattah & Ren, 2009; Kupiec, Pedersen, & Chen, 1995; Mendoza, Bonilla, Noguera, Cobos, & León, 2014). Paragraphs at the beginning and end of a document are more likely to

18

contain material that is useful for a summary, especially the first and last sentences of the paragraphs (Gupta & Lehal, 2010a; Teufel & Moens, 1997; Xie, Liu, & Lin, 2008).

iii. Title method

Important sentences normally contain words that are presented in the title and major headings of a document (Kupiec et al., 1995; Qazvinian, Hassanabadi, & Halavati, 2008; Shareghi & Hassanabadi, 2008). Thus, words occurring in the title are good candidates for document specific concepts (Teufel & Moens, 1997).

iv. Cue method

Cue phrases are words and phrases that directly signal the structure of a discourse. They are also known as discourse markers, discourse connectives, and discourse particles in computational linguistics (Hirschberg & Litman, 1993). Cue phrases, such as "as a *conclusion*" or "*in particular*" are often followed by important information. Thus, sentences that contain one or more of these cue phrases are considered more important than sentences without cue phrases (Zhang, Sun, & Zhou, 2005). These cue words are context dependent. However, due to the existence of different types of text, such as scientific articles and newspaper articles, it is difficult to collect these cue words as a unique list. Hence, since discourse markers can be used as an indicator of important content in a text and are more generic (Fraser, 1999), a list of cue words can be collected using discourse markers. Tables (A.1 to A.5) of Appendix A present the main discourse marker list (671 words). They are collected from previous studies (L Alonso, 2005; Fraser, 1999; Knott, 1996). In our work, in order to consider "*Cue method*", the list of cue words

extracted from these tables is presented in Table A.6. Although the produced list is not perfect, it can be used to identify cue method.

a. Cue Phrases: Linguistic Markers of Relations in RST

Rhetorical Structure Theory (RST) is a theory of text organization proposed in the 1980s as a result of exhaustive analyses of texts. It is a linguistically useful method for describing natural text, characterizing their structure primarily in term of relations that hold between parts of the text. It provides a way to explain the relations among clauses in a text, whether or not they are grammatically or lexically signalled.

RST was developed at the Information Sciences Institute of the University of Southern California by a group of researchers interested in Natural Language Generation. RST (Mann & Thompson, 1987, 1988) is based on the analyses of over several texts. The analysis is based on the assumption that some text units are more central (salient) to the text than others, and that the other units are given to support the reader's belief in them. The central units are named nuclei, and the supporting units are named satellites. Rhetorical relations are described in terms of schemas, i.e. the way in which one or more satellites (or nuclei) are related to the current nucleus.

The RST has been employed in a number of areas in discourse analysis, theoretical linguistics, psycholinguistics, and computational linguistics to plan coherent text and to parse the structure of texts. It can also be used to determine how coherence in text is achieved.

RST indicates text organization by means of relations that hold between parts of a text. It explains coherence by connected structure of texts, in which every part of a text has a role, a function to play, with respect to other parts in the text. The

relations have also been named coherence relations, discourse relations or conjunctive relations in the literature.

Coherence relation is the property of well-written texts that makes them meaningful, easier to read and understand than a sequence of randomly string sentences (Lin et al., 2011). Coherence relation between sentences is considered as keys for the ability to understand or generate discourse. This is because sentences are not generally understood in isolation, but with respect to others (Lascarides & Asher, 1993; Maier & Hovy, 1993; Mann & Thompson, 1988; Marcu & Echihabi, 2002; Martin, 1992). Coherence relations are categorized into two types: explicit relations and implicit relations. Explicit coherence relations are signaled by cue phrases that point to them. In contrary, implicit coherence relations can only be detected from the context, and syntax of the discourse itself, as well as from the knowledge domain of the text (Taboada, 2009). Often, discourse coherence relations are explicit, by the use of appropriate cue phrases such as the cue phrase "because" in the following example,

Example 1: "*I am very sad **because** I lost my book.*"

The example includes two sentences related together by a "causality" relation, and the cue phrase "because" which explicitly connects them by the "causality" relation.

However, when the discourse relation is implicit, it could be determined from the context and syntax of the discourse itself, as well as from the knowledge domain of the text. If the text in example 1 is reformed again without the connector "because", the same "causality" relation still remains but in an implicit form, as shown in example 2.

Example 2: "*I am very sad. I lost my book.*"

The example 2 display the text that includes two sentences and the cue phrase "because" is absent but the coherence relations can be guesstimated from the context. Such kinds of relations, as the example 2, are so called implicit, un-signaled, or hidden coherence relation. In this case, the RST recognizes relations that are, seemingly, not signalled in any explicit way.

Mann and Thompson (1988) introduced 24 relations, which can be grouped into subject matter (e.g. Elaboration, Circumstance, Solution hood, Cause, Restatement) and presentational relations (Motivation, Background, Justify, Concession). Presentational relations are those whose intended effect is to increase some tendency in the reader, such as the desire to act or the degree of positive regard for, belief in, or acceptance of the nucleus. Subject matter relations are those whose intended effect is that the reader recognizes the relation in question. Each group includes relations that share a number of characteristics and differ in one or two particular attributes. The relation definition does not rely on morphological or syntactic signals. The relation always determine based on functional and semantic judgements.

Table 2.1 shows sample of the defined relations, *N* stands for nucleus, *S* for satellite, *W* for writer and *R* for reader.

Definitions of Relations			
Relation Name	**Constraints on either** S **or** N **individually**	**Constraints on** $N+S$	**Intention of** W
Evidence	On N: R might not believe N to a degree satisfactory to W. On S: R believe S or will find it credible.	R's comprehending S increases R's belief of N	R's belief of N is increased
Justify	None	R's comprehending S increases R's readiness to accept W's right to present N	R's readiness to accept W's right to present N is increased
Motivation	on N: N is an action in which R is the actor (including accepting an offer), unrealized with respect to the context of N	Comprehending S increases R's desire to perform action in N	R's desire to perform action in N is increased
Condition	On S: S presents a hypothetical, future, or otherwise unrealized situation (relative to the situational context of S)	Realization of N depends on realization of S	R recognizes how the realization of N depends on the realization of S

- **Invention**

The invention rule is used when there are no explicit topic sentences in paragraphs. In such cases, one should make up explicit topic sentences by using his or her own words to state the implicit main idea of paragraphs. Thus, the invention rule requires that students "add information rather than just delete, select or manipulate sentences already provided for them" (Brown & Day, 1983).

- **Copy-verbatim**

In the copy-verbatim process, a summary sentence is produced from the source sentence without any changes. This strategy is not part of the summarizing strategies but it is used by students.

2.2.2. Current systems to identify summarizing strategies

A few systems have been proposed to identify summarizing strategies. To the best of my knowledge only two systems were proposed to identify summarizing strategies. In this subsection, we discuss about these systems in detail:

Modelling summarization assessment strategies (MSAS) (Lemaire et al., 2005) based on LSA have been developed where using LSA, the summary text is semantically compared with the source text to identify the summarizing strategies, including copy-verbatim, paraphrase, construction and generalization. LSA has some disadvantages, the first of which is that it does not use syntactic composition, such as word order in comparing two sentences. The second limitation is that it can produce a reasonable result when it takes a large corpus as its input but is not suitable for short text. The third limitation is that since not all of the words appear in all the sentences, the created matrix is usually sparse. Finally, most of the models that are based on LSA use a similarity threshold to make a decision; however, determining the value of the threshold is difficult.

Summary Sentence Decomposition Algorithm (SSDA) (Idris et al., 2009), which is based on word position, has been proposed to identify the summarizing strategies used by students in summary writing. Using a syntactic composition (word position), the summary text is syntactically compared with the source text to identify the summarizing strategies, including deletion, sentence combination, syntactic transformation, sentence reordering and copy-verbatim. It does not use the semantic relationships between words in comparison to sentences, and hence, it cannot find summarizing strategies at the semantic level, such as paraphrasing, generalization, and invention.

2.3 Text Summarization Systems

The advancement in electronically available documents makes research and applications in automatic text summarization more significant. However, the huge number of available in digital media makes it difficult to obtain the necessary information related to the needs of a user. To solve this issue, text summarization systems (TSS) can be used. .

Text summarization systems produce a summary of one or more text automatically. The summary normally contains the aim, approaches, results, and conclusions presented in the source text and remove needless words, phrases, and sentences. The purpose of automatic summarization is to produce a summary from a source by extracting the important content from the source text and display it to the user in a compressed form (Saggion & Poibeau, 2013). By using the summary produced, a user can decide if a document is related to his or her needs without reading the whole document.

2.3.1. Phases of Text Summarization Systems

In general, summarization can be divided into 3 steps (Laura Alonso et al., 2004; Gholamrezazadeh, Salehi, & Gholamzadeh, 2009; K Sparck Jones, 1999; Lloret, 2012), which are:

- *Interpretation:*

  The input document is exposed in a format that the processing can be performed on it.

- *Transformation:*

  Input presentation is reformed into summary presentation.

- *Generation:*

Summary presentation is changed into summary text.

2.3.2. Important aspects of Text Summarization Systems

Figure 2.1 present three important aspects of text summarization. These aspects include input aspect, purpose aspect and output aspect (Alemany, de Lingüıstica General, Masalles, & Cirera, 2005; K Sparck Jones, 1999; Mishra et al., 2014). We describe each of them as follows.

i. Input aspects

The characteristics of input text can affect the result of summary, according to the following aspects:

- *Document configuration:* different information can be found in the source text. For example, labels those show headers, chapters, section, lists and tables.

- *Domain*: The input source text can be connected to a specific topic, or can be general.

- *Language*: system may be language related or non-language related.

- *Unit*: the input to the text summarization can be a single document, multi document, and multimedia information.

- *Scale*: different summarizing strategy has to handle various text lengths.

ii. Purpose aspects

Summarization systems can produce summaries of a given source text. The following factors are related to the purpose aspects of summarization systems.

- *Situation*: The environment that the summary will be used; in other word who uses the summary.

- *Audience:* The reader, who reads the summary.

- *Use: T*he purpose for creating the summary.

iii. Output aspects

The result of the summary can be affected by the following output aspects:

- *Content:* a summary can consist of all aspects and main concept of a source text or it may focus on some specific aspects which are determined by a query.

- *Format*: a summary can be a simple text, or it can be organized by header or tags.

- *Style*: A summary can be informative, indicative, aggregative, or critical. Informative summaries cover the topics of the source text. Indicative summary produces a concise survey of topics that are mentioned in the original text. Aggregative summaries provide extra information that does not exist in the input text. Critical summaries check true and false elements of the input document.
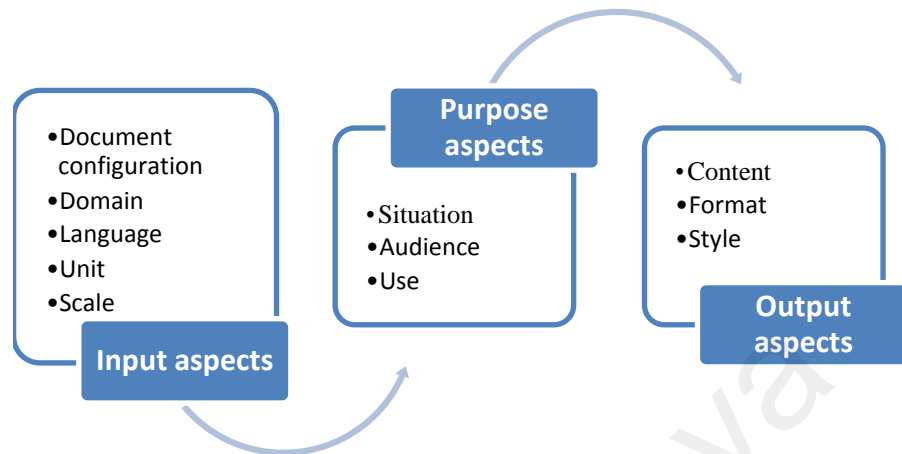
Figure 2.1: Phases of Text Summarization Systems

2.3.3. Categorization of Text Summarization Systems

Figure 2.2 presents the categories of text summarization systems. The output of the system may be an extractive or abstractive summarization. An *extractive summarization* method comprises of selecting important sentences from the original text (Gupta & Lehal, 2010b). The importance of sentences is determined by statistical and linguistic features of sentences. An *abstractive summarization* (Erkan & Radev, 2004; Hahn & Romacker, 2001) tries to develop a comprehension of the main concepts in a text and then expose those concepts. It uses linguistic methods to analyse and interpret the text and then to find the new concepts and expressions to best describe it by generating a new concise text that takes the most important information from the original text.

A summarization system can be based on *single or multiple documents* (Goldstein, Mittal, Carbonell, & Kantrowitz, 2000; Hovy & Lin, 1998; Mendoza et al., 2014). In single document summarization system, a single-document is used to generate a summary, while in multi-document summarization systems, multiple documents on the same subject are used to generate a single summary. Besides these facts, text

summarization system can also be either *indicative or informative summarization.* Indicative summarization systems only present the main idea of the text to user. The typical length of this type of summarization is between 5 to 10 per cent of the main text. Indicative summaries can be used to encourage the readers to read the main documents (Hovy & Marcu, 2005). The informative summarization systems give concise information of the main text and it can be considered as a substitution for the main document. The length of informative summary is between 20 to 30 per cent of the main text (AlSanie, Touir, & Mathkour, 2005).

Summarization systems can also be categorized into generic and query-based summarization systems. In generic text summarization, the summary is made about the whole document. However, in query-based text summarization, the provided summary is based only on the specific query (Jing & McKeown, 2000; Sarker, Mollá, & Paris, 2013).
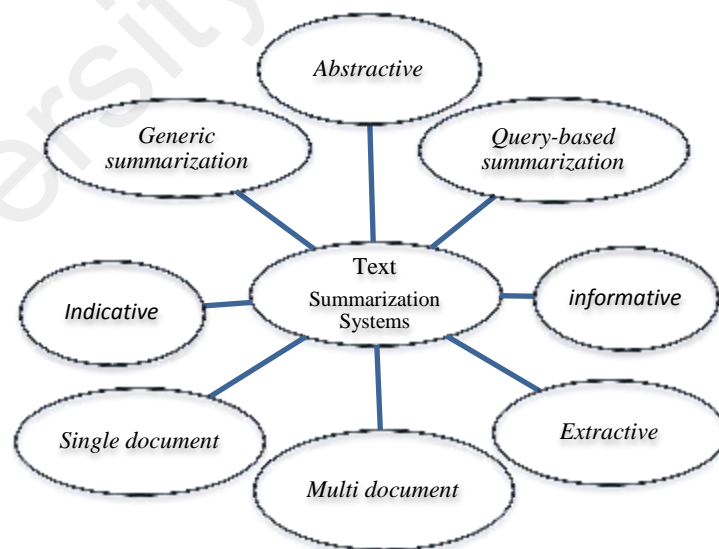
Figure 2.2: Text summarization Categorization

2.3.4. Approaches to text summarization systems

There are many different approaches to text summarization in literature. This section explains some of these approaches.

i. Surface Level Approaches

The oldest approaches use surface level indicators or shallow features to identify important sentences of a document. These features include word frequency, sentence location, title word and cue words or phrases.

Luhn (1958), Ferreira et al. (2013), Alguliev, Aliguliyev, and Isazade (2013), Cai and Li (2011), Wang and Li (2012) and Glavaš and Šnajder (2014) used the term frequency technique to produce a summary of a document. The idea was that more frequent words are most important. The sentences that include these frequent words are assumed to be more important than other sentences, and are selected to be part of the summary text. However, not all the words in the document are taken into consideration. For example, stop words are not used for calculating the term frequency. Keywords are usually nouns and verbs. Key word is determined using $tf \times idf$ measure. The term frequency TF value is the number of occurrences of the term in a document. The inverse document frequency IDF value is calculated using the equation (2.1):

$$\text{IDF} = log \frac{|N|}{n_i} \tag{2.1}$$

Where,

$|N|$: is the total number of document in the input text.

n_i : is the number of document that contains the term.

The location of sentences can give information about the importance of that sentence. Usually, the first and the last sentence of the first and the last paragraph of a text

30

document are more important and they have greater chances to be included in summary. The algorithms belonging to Baxendale (1958), Edmundson (1969) and Brandow, Mitze, and Rau (1995) are examples to the approaches that use position of words or sentences.

Title word feature assumes that the important sentences normally contain words that are presented in the title or headings (Gupta & Lehal, 2010b; Kupiec et al., 1995; Teufel & Moens, 1997).

Cue phrases are words and phrases that directly signal the structure of a discourse. These words are also known as clue words, discourse markers, discourse connectives, and discourse particles in the computational linguistic (Hirschberg & Litman, 1993). Cue phrases can be defined as a set of lexical signals that make coherence relations explicit in the surface (Hirschberg & Litman, 1993). Cue phrases such as "as a *conclusion*" and "*in particular*" are often followed by important information. Thus, sentences that contain one or more of these cue phrases are considered more important than sentences without cue phrases (Zhang et al., 2005). Cue phrases can be defined as a set of lexical signals that make coherence relations explicit in the surface text, including connectives, clause conjunctions, subordinators and sentential adverbials (Fraser, 1999). These lexical expressions are considered under different name, such as, discourse markers, discourse connectives, discourse operators, pragmatic connectives, sentence connectives, and cue phrases. Lexical expressions are classified into three syntactic classes such as conjunctions (*but, and, or*), adverbs ("*consequently*", "*conversely*," *equally*"), and prepositional phrases ("*as a consequence*," *in particular*", "*after all*," *on the other hand*").

These lexical expressions are also classified into three main classes (Fraser, 1999) as:

- *Contrastive,* interpretation of *Sentence* 2 contrasts with an interpretation of *Sentence*1 (*such as, (al) though, but, contrary to this/that, conversely, despite (doing) this/that, however, nevertheless*).

    Example: "*We left late. <u>Nevertheless</u>, we got there on time.*"

- *Collateral,* a parallel relationship between *Sentence* 2 and *Sentence* 1 (*such as above all, also, analogously, and, besides, better yet, by the same token, correspondingly*).

    Example: "*You should always be polite. <u>Above all</u>, you shouldn't belch at the table.*"

- *Inferential, Sentence* 2 shows a conclusion based on *Sentence* 1 (*accordingly, all things considered, as a (logical) consequence/conclusion, as a result, because of this/that, consequently*).

    Example: "*It's raining. <u>Under those conditions</u>, we should ride our bikes.*"

ii. Graph Based Approaches

Graph theoretic (Kruengkrai & Jaruskulchai, 2003) provides a method to identify the important sentences of a document. After the common pre-processing steps, namely, stop word removal and stemming, each sentence of a document is represented as a node in graph, in other word, there is a node for every sentence. The edges represent the similarity among the sentences. If two nodes (sentences) share some common words, they will be connected with an edge. Then, once the networking is built, the system finds the important sentences. In this case, the sentences with highest similarity to the other sentences are chosen as a part of the resulting summary. Text Rank (Mihalcea &

Tarau, 2004) and Cluster Lex Rank (Qazvinian & Radev, 2008) are two methods that use graph based approach for document summarization.

iii. Query based approaches

In query based text summarization (Pembe & Güngör, 2007), each sentence of a given document is scored based on the frequency counts of terms (words or phrases). The sentences containing the query phrases are given higher scores than the ones containing single query words. Then, the sentences with highest scores are combined into the output summary text. Sentences of text may be extracted from different sections or subsections. The resulting summary includes sentences that can be extracted from different sections. The number of sentences depends on the summary frame size which is inputted as a variable by the user.

2.4 Summarization assessment

2.4.1. What Is Assessment?

Assessment comprises the use of experimental data on student learning to clarify programs and enhance student learning (M. Allen, 2004). It is the process of collecting information from several sources in order to expand a deep conception of what students know, understand, and can do with their knowledge as a result of their educational experiments (Huba & Freed, 2000). Assessment is the systematic base for creating inferences about the learning and progress of students. It is the process of specifying, review, choosing, planning, gathering, analysing, understanding, and using information to enhance students' learning and progress (Erwin, 1991; Palomba & Banta, 1999).

There are various types of assessment used in practice such as: multiple-choice question, true-false question, short answer question, essay writing and summary writing

(Zipitria et al., 2010). Multiple-choice questions ask students to select one choice from a given list. The correct answer in the list is – called the key, and the rest – the several wrong answers – are called districts. Multiple-choice questions are used for measuring comprehension. For short answers, the required answer may be a word or phrase, one sentence or more. A student needs to prepare the appropriate words, numbers, or symbols to answer a question or complete a statement. True-false questions are usually a declarative statement that the student must judge as true or false. The essay is used to determine what students know about a few topics. Essay consists of a number of questions which the students display their abilities in recall knowledge and organize this knowledge into a writing product.

Summary writing can help students to develop their ability to read and write (Graham & Hebert, 2010; Westby et al., 2010). In fact, a summary is used to organize and emphasize the most relevant content of the text. Previous studies have shown that it is effective to improve comprehension (Chiu et al., 2013; Hartley & Trueman, 1983; Kintsch et al., 2000; Jose A León & Carretero, 1995). Some researchers have suggested that if readers are not capable to summarize a text, then they have not understood it (Palinscar & Brown, 1984). Moreover, a summary is a brief expression of the most important information in a text and describes most of the main ideas in a text. Other than that, summary is an effective resource of creating and integrating new knowledge (José A León, Olmos, Escudero, Cañas, & Salmerón, 2006).

2.4.2. Types of Summarization Evaluation

There are two types of summarization assessments: intrinsic and extrinsic. The two types of intrinsic summarization evaluations are human and automatic. Human intrinsic evaluations assess the summarization system itself, based on factors such as clarity, coherence, fluency and informativeness (Jing, Barzilay, McKeown, & Elhadad, 1998). These will be discussed below in Section 2.4.3. Automatic intrinsic evaluation measures

usually compare a candidate summary (output of a summarizer) against a model human summary (Mani et al., 2002). These will be discussed in Section 2.4.4.

Extrinsic evaluations study the use of summarization for a specific task. An extrinsic evaluation tests the summarization based on how it affects the completion of some other tasks. Examples of such task are: (1) information retrieval, (2) question and answering (Inderjeet, 2001).

2.4.3. Human Intrinsic Assessment

In human intrinsic assessment, two major factors such as writing style and informativeness of a summary or content of a summary are considered (Farzindar, Rozon, & Lapalme, 2005; Mani et al., 2002). Each of these factors is explained as follows:

- *Content*

    The extracted summary (student-written summary) is compared with ideal summaries (prepared by teachers). The assessment should measure not just the amount of the content that students know, but the extent of their understanding, their ability to make connections, and their ease in transferring knowledge to new situations.

- *Writing style*

    Writing style of a summary can be evaluated by assessing the readability based on different criteria such as:

    - Use *accurate* spelling, punctuation and grammar.

    - Communicate *creatively,* using a varied range of vocabulary, sentence structures and linguistic devices.

    - Communicate *clearly* and develop ideas coherently, at word level, at sentence level and at whole text level.

- Communicate ***appropriately***, with a clear awareness of purpose and audience.

2.4.4. Automatic Intrinsic Assessment

Automatic intrinsic summarization assessment usually compares a candidate summary with a model summary. Various techniques such as BLEU, ROUGE, LSA and N-gram, and numerous systems based on these techniques have been proposed for summarization evaluation. This section reviews the current techniques and the summary assessment systems based on the existing techniques.

Latent Semantic Analysis (LSA) (Landauer, 2002) is a statistical technique for representation of the meaning of words and sentences. It has been used in educational applications, such as essay grading (Landauer & Dumais, 1997), as well as in NLP applications such as information retrieval (Landauer, Laham, Rehder, & Schreiner, 1997) and text segmentation (Choi, Wiemer-Hastings, & Moore, 2001). It only requires raw text as its input and represents the text as a matrix in which each row represents a unique word and each column represents a text passage or sentence. Each cell is used to represent the importance of words in sentences. Different approaches can be used to fill out the cell values, such as frequency of words, Binary Representation and Term Frequency-Inverse Document Frequency. Then Singular Value Decomposition (SVD) is applied to the matrix. Although LSA is used in several applications, it has some disadvantages: 1) it does not use syntactic composition, such as word order, in which this information is necessary to understand the meaning of two sentences; 2) it can produce a reasonable result when it takes a large corpus as its inputs; 3) since all words do not appear in all sentences, the matrix created is usually sparse.

Bilingual Language Evaluation Understudy (BLEU) is an n-gram precision based evaluation metric initially designed for the task of machine translation evaluation (Papineni, Roukos, Ward, & Zhu, 2002). The developers of BLEU also suggest that this metric could be used for summarization evaluation where the metric can determine how much the information in a summary text overlaps with the information in a source text using N-gram co-occurrence statistics. BLEU's n-gram precision is defined as:

$$P_N = \frac{\sum_{C \in \{Candidates\}} \sum_{N-gram \, \varepsilon \, C} Count_{clip}(N-gram)}{\sum_{C \in \{Candidates\}} \sum_{N-gram \, \varepsilon \, C} Count(N-gram)} \tag{2.2}$$

Where *Count_{clip} (N-gram)* is the maximum number of N-grams co-occurring in a candidate translation and a reference translation, and *Count (N-gram)* is the number of N-grams in the candidate translation. The equation is known as precision because the denominator is the total number of n-grams in the candidate translation. BLEU also imposes a *brevity penalty* (BP) to ensure that short candidate translations are not unfairly scored very highly. Let c be the length of the candidate translation and r be the reference translation length. The BP is calculated as:

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{\left(1-\frac{r}{c}\right)} & \text{if } c \le r \end{cases} \tag{2.3}$$

$$BLEU = BP \times e^{\left(\sum_{n=1}^{N} W_n \times \log P_n\right)} \tag{2.4}$$

Where $N \ge n$, n and N range from 1 to 4, $W_n = 1/N$.

Researchers at the University of Southern California's Information Sciences Institute (ISI) proposed a new recall-based evaluation metric, Recall Oriented Understudy of Gisting Evaluation (ROUGE). It includes several automatic evaluation methods that measure the similarity between summaries (Lin, 2004; Lin & Hovy, 2003).

- ROUGE-N: N-gram Co-Occurrence

ROUGE-N, is an n-gram recall between a candidate summary and a reference summary. ROUGE scoring is computed as:

$$\text{ROUGE} - \text{N} = \frac{\sum_{C \,\epsilon\, \{\text{Reference Units}\}} \sum_{\text{N-gram} \,\epsilon\, C} \text{Count}_{\text{match}}\ (\text{N-gram})}{\sum_{C \,\epsilon\, \{\text{Reference Units}\}} \sum_{\text{N-gram} \,\epsilon\, C} \text{Count}\,(\text{N-gram})} \qquad (2.5)$$

Where *Count$_{match}$ (n-gram)* is the number of n-grams co-occurring in a candidate summary and a reference summary, and *Count (n-gram)* is the number of n-grams in the reference summary. This equation is recall-based because the denominator is the total number of n-grams in the reference summaries.

- ROUGE-L: calculates the similarity between a reference summary and a candidate summary based on the Longest Common Subsequence (LCS).
- ROUGE-S: a measure of the overlap of skip-bigrams between a candidate and a reference summary.

Summary Street (Franzke & Streeter, 2006), which is based on LSA is a computer-based assessment system that is used to evaluate the content of the summary text. Summary Street ranks a student's written summary by comparing the summary text and the source text. It creates an environment to give appropriate feedback to the students, such as content coverage, length, redundancy and plagiarism.

A commercial summary evaluation system, Laburpen Ebaluaka Automatikoa (LEA) (Zipitria et al., 2004), which is based on Latent Semantic Analysis (LSA), has been proposed to evaluate the summary. It is designed for both teachers and students. It allows teachers to examine a student's written summary, and allows students to produce a summary text with their own words. The summaries are evaluated based on certain

features, such as cohesion, coherence, use of language and the adequacy of the summary.

Evaluation measures for text summarization presents a new evaluation measure for assessing the quality of a summary (Steinberger & Ježek, 2012). The core of the measure is LSA. The Automatic Assessment of Students' free-text answers (Pérez, Alfonseca, & Rodrıguez, 2004) is based on the Bilingual Evaluation Understudy (BLEU) algorithm and LSA was developed for grading students' essays. The system compares a student's essay and the model essay to determine how similar they are.

Lin and Hovy (2003) proposed a system based on BLEU and N-gram co-occurrence to evaluate summaries with the aim of measuring the closeness of the summary text to the source text. Y. He et al. (2009) proposed a summary assessment system based on the modified LSA algorithm and N-gram co-occurrence with the aim of assessing students' written summaries.

Framework for Evaluating Summaries Automatically (FRESA) (Fan, 2010a), which is based on Jensen-Shannon divergence and ROUGE is a framework that is used to evaluate the multilingual summarization without Human references. It used the Rouge package (Lin, 2004) such as uni-grams, bigrams, and the skip bi-grams with maximum skip distance of 4 (ROUGE-1, ROUGE-2 and ROUGE-SU4), to compute various statistics. The Jensen-Shannon divergence formula is calculated as follows:

$$D_{JS}(P||Q) = \frac{1}{2} \sum_w P_w \, log \, 2 \, \frac{2P_w}{P_w + Q_w} + Q_w \, log \, 2 \, \frac{2Q_w}{P_w + Q_w} \qquad (2.6)$$

Where P is the probability distribution of words w in text T and Q is the probability distribution of words w in summary S.

Nouveau-ROUGE (Conroy, Schlesinger, & O'Leary, 2011) is based on the ROUGE (Recall-Oriented Understudy for Gisting Evaluation), and was developed for updating

summarization. It used a ROUGE score to measure similarity, and thus redundancy, between a given original summary and an update summary. Metrics to predict the quality of a Text Summary (Rankel, Conroy, & Schlesinger, 2012), which is based on the several features, is a model to evaluate the quality of a summary. The model combines linguistic features such as the number of sentences, the measure of redundancy and sentence length and content feature such as ROUGE (ROUGE-SU4,ROUGE-2) for assessing the linguistic and content quality.

Mohler, Bunescu, and Mihalcea (2011) introduced an Answer Grading System, which combines a graph alignment model and a text similarity model. This system aims to improve the existing approaches that automatically assign a grade to an answer provided by a student, using the dependency parse structure of a text and machine learning techniques. The current system uses the Stanford Dependency Parser (De Marneffe, MacCartney, & Manning, 2006) to create the dependency graphs for both the student (A_1) and teacher (A_2) answers. For each node in the student's dependency graph the system computes a similarity score for each node in the teacher's dependency graph using a set of lexical, semantic, and syntactic features. The similarity scores are used to weight the edges that connect the nodes in A_1 on one side and the nodes in A_2 on the other. The system then applies the Hungarian algorithm to determine both an optimal matching and the score associated with such a matching for the answer pair. Finally, the system produces a total grade based on the alignment scores and semantic similarity measures.

2.5 Summary

In this chapter, we reviewed some important topics in summarization. We introduced summarization by describing the definition of summary, classification of summary, phases of text summarization systems, investigation on summary aspects, efficacy of summarization on comprehension, summarizing strategies in summarization, summarization assessment, intrinsic and extrinsic evaluations in text summarization. Evidently, summarization has been widely studied, but little work has been done on summarization assessment. Some of the current techniques and summary assessment systems based on the existing techniques were discussed. Most research into summarization assessment is very much focused on the output of the summary writing process, especially on the content. A few systems have been proposed to identify summarizing strategies.

CHAPTER 3

DIRECT INSTRUCTION ON SUMMARY WRITING

3.1 Introduction

Summary writing is a process to determine the important information of the text and explaining these ideas in our own words. In summary writing the vocabulary and the sentence structure of a source text will be changed, but the main ideas remain. The main objective of summary writing operation is to produce a summary text from source text.

A summary text includes only main points and main idea from the source text. It does not include repetitive information. In addition, it is also shorter than the source text and takes a wide overview of the source text. Writing a good summary demonstrates that the students clearly understand the text. It also helps them to understand and remember information they read.

3.2 Summary writing as Teaching Tool

Most researchers have found summary writing to be effective in improving both reading comprehension (Chiu et al., 2013; Westby et al., 2010) and writing in English as a Second Language (ESL) classes (Alyousef, 2006; Cho, 2012; Fan, 2010b; Hedge, 2001; Pakzadian & Rasekh, 2013). They argue that summarizing skills are essential to academic success and should be taught in schools. In addition, summary writing improves not only student's reading and writing, but, also their vocabulary since students need to use new words in summaries by paraphrasing some words. Finally, the critical thinking skills will also be improved as students have to decide on the main ideas of the text to be included in the summary (Koda, 2005).

According to the previous studies, summary writing does not develop by trial and error, but it needs instruction (Westby et al., 2010). When told to summarize, students will often either copy verbatim, write long summaries, or write very short ones until losing most of the key points (Taylor, 1983). This happens because students do not fully comprehend what a summary is or how to write it. However, summarizing is a technical skill (Sobh et al., 2006; Yang, Chen, et al., 2013). Thus, the aim of summary writing instruction is to focus on how to arrange information, how to certify that the summary is not changing the main idea of the source text, what to remove and what to contain.

3.2.1. Steps in Summary writing

The following steps as shown in Figure 3.1 (1), are taught to students for writing a summary (Riedl, 2011; Swales, 1994). In each step, we explain how each step of training summary writing corresponds to each summarizing rule, as suggested by these authors (Brown & Day, 1983; Idris et al., 2009; Kintsch & Van Dijk, 1978; Lemaire et al., 2005), as shown in Figure 3.1 (2). These steps are described in detail as follows:

Step1. Find the important information from the paragraphs

The first step is to read and focus on any headings, subheadings, first and last paragraphs and first and last sentence of the following paragraphs. It is also suggested to underline any bold-faced terms, the sentences from the text that are related to the heading or subheading of a paragraph and identify the details that should be avoided. Finally, to find the important sentences that convey the main idea of the paragraphs, where the sentences normally contain key point and not minor details.
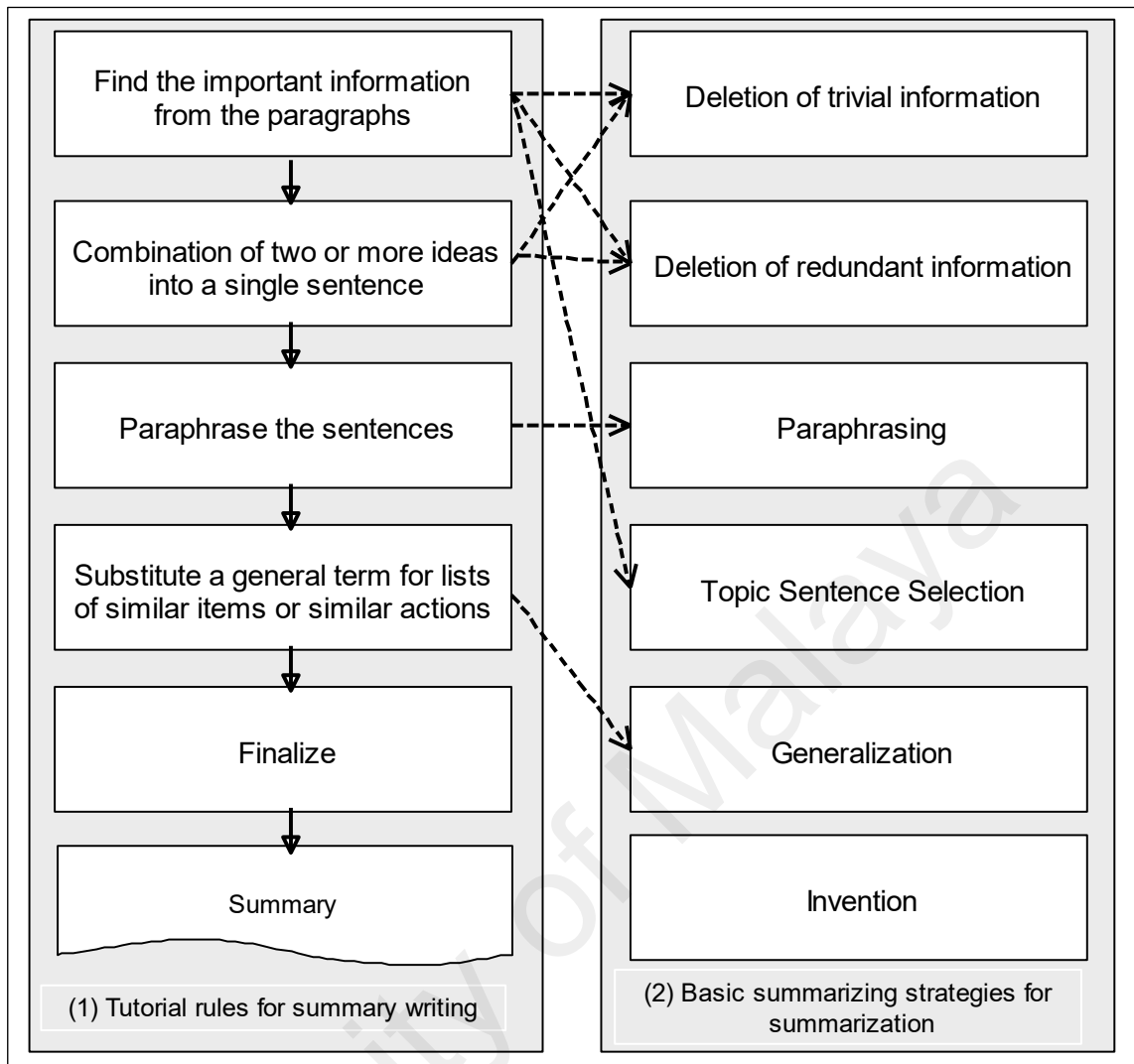
Figure 3.1: Steps in summary writing and basic summarizing strategies for

summarization

The first step of teaching summary writing uses some of the basic summarizing rules

such as deletion of trivial information, deletion of redundant information and topic

sentence selection. There are four methods in topic sentence selection rule that are used

to identify important sentences from paragraphs. These methods are:

1. Title method– important sentences include words that are presented in the

   headings or sub headings of the paragraphs.

2. Cue method– words or phrases such as "*in summary*", "*in conclusion*" are often

   used to indicate the importance of the sentences. Thus, sentences that include

these words or phrases are considered more important than sentences without these phrases.

3. Key method– the words with high frequency in the text are most representative of its content, thus a sentence which comprises these words more relevant.

4. Location method– important sentences are normally at the beginning and the end of a document or paragraph.

Step2. *Combination of two or more ideas into a single sentence*

In this step, students are taught on how to merge two or more sentences from the original text or two or more phrases of sentences from the original text into a single sentence in the summary text. This step uses deletion rules to combine two or more main idea into a single sentence. This strategy also uses conjunction words to connect ideas, paragraphs and sentences. These conjunction words are also classified in some main classes (Fraser, 1999; Swan, 2005; Willis, 1981) as:

i. Coordinate Conjunctions

Coordinate conjunctions are used to connect an idea between two clauses (a clause could be a part of a sentence or a full sentence in itself) or sentences. The most common coordinating conjunctions are: *and, or, nor, but, yet,* and *so.*

Example: "*Sometimes the demand goes up* <u>and</u> *the price does not fall.*"

ii. Correlative Conjunctions

Correlative conjunctions are used to connect two parallel phrases or clauses. Some of correlative conjunctions words are: *both…and; not only…but also; just as…so*.

Example: "<u>Not only</u> *did I respect his intelligence,* <u>but</u> *I* <u>also</u> *admired his qualifications."*

iii. Subordinate Conjunctions

Subordinate Conjunctions create one clause dependent on, or subordinate to, the other. Example of Subordinate conjunctions words are: *although, even though, whereas, while, before, after, until, when, while, since.*

Example 1: "<u>Although</u> *we all know smoking is bad for our health, many people still smoke."*

Example 2: "*The students passed their exams* <u>because</u> *they had studied hard."*

Step3. Substitute a general term for lists of similar items or similar actions

In some cases, we need to replace a list of words with a general word which belongs to the same class. Hence, this step corresponds to the generalization strategy. The aim of this strategy is to replace a list of words by a more general term. There are two types of replacements. One is the replacement of a general word for a list of similar item. The other one is the replacement of a general term for a list of similar actions.

Step 4.Paraphrase the sentences

Paraphrasing is a way to replace a word from a sentence of source text with a similar word or synonym (words that share the same meaning) in summary sentence. Hence, this step corresponds to the paraphrase strategy. This strategy is a good skill, because it helps students to expand their vocabularies.

Guidelines for effective paraphrasing:

1. Read the sentence until you understand its meaning.

2. Write your paraphrase.

3. Check the new sentence with information from the source sentence to make sure the produced sentence represents similar meaning of the source sentence.

Step 5.Finalize

Read your summary to make sure that your idea is clear to someone who did not read the source text. On the other hand, read your summary to look for mistakes such as grammar, verb tense, spelling and punctuation errors.

As displayed in Figure 3.1, invention strategy is not part of the steps in tutorial rules (Figure 3.1(1)). The invention strategy is employed when there is no explicit sentence in a paragraph, which represents the main idea of paragraph. In this case, a summarizer should produce explicit topic sentences by using his or her own words to state the implicit main idea of the paragraphs. Thus, the invention strategy is regarded as difficult and the skill of using this strategy will improve with age and experiences (Brown & Day, 1983). Due to its difficulty, it is not taught in school.

A walk-through example of the summarizing rules is presented in the following section:

3.2.2. Summary writing Strategies

Summarizing involves a set of conscious process to extract the main ideas from the source text. These processes are some kind of rules that link textual proposition from different paragraph of the source text for making the summary text. These rules are called summarizing strategies or summarizing rules. Some of these summarizing

strategies are used to eliminate the unnecessary information, whereas, some of these strategies are used to combine important information to produce new information. In particular, there is a difference between these summarizing rules. We describe some of these summarizing rules with an example as follows.

Deletion – this rule eliminates all unnecessary details from the text that do not contribute to extract the main idea of the text. It also removes words and propositions that repeat information. In other words, it omits redundant information from the text. The following example is used to explain the deletion rule.

Source text:

> *"The newspaper today also motivates students to learn because it contains educational supplements that help them with their lessons. "*

Summary text:

> *"The newspaper motivates students to learn."*

Topic Sentence Selection (TSS) – There are four methods which are title method, cue method, key method and location method, to identify the important sentence. The following examples are used to explain the topic sentence selection rule.

1. Location method (The first or last sentence of a paragraph)

Source text:

> *"Text Summarization is the process of identifying the most salient information in a document or set of documents. Summarization shares some basic techniques with indexing as both are concerned with identification of the essence of a document. Also, high quality summarization requires sophisticated NLP techniques in order to deal with various Parts Of Speech (POS) taxonomy and inherent subjectivity."*

Summary text:

> *"Text Summarization is the process of identifying the most salient information in a document or set of documents."*

2. Cue method

Source text:

> *"In one experiment, 84 low-income, minority high school juniors were randomly assigned to teachers who implemented direct instruction of summary skills (inductive instruction) or a control group of teachers did not provide explicit instruction on summarization strategies (deductive instruction). <u>In conclusion,</u> the findings indicate that direct instruction of*

summarization skills has a significant effect on improving the skills of students."

Summary text:

"Findings indicate that direct instruction of summarization skills has a significant effect on improving the skills of students."

3. Title method

Source text:

Title: *"try to get the boat from the water"*

Paragraph:

"While my father was watching me, he suddenly remembered that he needed to take his medicine. He instructed me to come back to the shore for a moment before he rushed into the staff kitchen to take his pill. I started towards the shore as I saw my father turned away. But when I reached out to get my boat, a soft wind caught its tail and pushed it away from me."

Summary text:

"When I reached out to get my boat, a soft wind caught its tail."

4. Key method

Suppose two words "*dive*" and "*water*" with high frequency appeared in the source text, therefore due to the key method, each sentence from a paragraph can be selected as topic sentence if the sentence includes the word with high frequency. Hence, from the following paragraph the sentence "*He repeatedly dived under the water*" can be chosen as a topic sentence that summarizes a whole paragraph.

Source text:

> *"When my father emerged from the hut, he was horrified to see me struggling in the water some distance from the shore. He dived in and swam as hard as he could to the spot where I had gone under. He repeatedly <u>dived</u> under the <u>water</u>, frantically searching for my body."*

Summary text:

> *"He repeatedly dived under the water."*

Sentence combination – in sentence combination strategy two or more sentences are usually combined using conjunction words, such as "*for", "but", "and", "after", "since"* and "*before"*. The following example is used to explain sentence combination rule.

Source text:

> *"A survey carried out on the benefits of the newspaper*
>
> *clearly shows that students who read the newspapers have*
>
> *better awareness of the world around them. This is because"*
>
> *they are better informed of the current issues happening*
>
> *around the world. Hence, they become more*
>
> *knowledgeable."*

Summary text:

> *"A survey carried out on the benefits of the newspaper*
>
> *clearly shows that students who read the newspapers have*
>
> *better awareness of the world **and** they are better informed*
>
> *of the current issues happening around the world."*

Generalization – deletion rule allows for the elimination of unnecessary information, whereas the generalization rule is to replace a general term for a list of words. There are two replacements: 1) the replacement of a general word for a list of similar items, e.g. "*hoopoe, Pigeons, hawk and vulture*" can be replaced by "*bird*"; 2) the replacement of a general word for a list of similar actions. The following example is used to explain the generalization rule.

Source text:

> a) *"Peter saw a hawk."*
> b) *"Harry saw a vulture."*
> c) *"John saw a hoopoe."*
> d) *"Yang saw a Pigeons."*

Summary text:

"The boys saw a bird."

Paraphrasing – by applying paraphrase strategy, a word from source sentence is replaced with a similar word or synonym in summary sentence. The following example is used to explain the paraphrase rule.

Source text:

"I <u>dived</u> and swam back to shore."

Summary text:

"I <u>plunged</u> into the ocean and swam back to shore."

In the previous chapter (Chapter 2) we have discussed five basic summarizing strategies which are sentence combination, deletion, paraphrase, topic sentence selection and invention. However, it is found that the invention strategy was not taught to students. According to the definition of invention strategy, often a summarizer produces a sentence that summarizes a whole paragraph. The summarizer should produce the sentence in his or her own words. Obviously, this strategy is difficult and develops along with growing older (Brown & Day, 1983). Hence, it can be one of the important reasons why this strategy was not taught in school. An example of invention strategy is as follows.

Source text:

a) *"Mary was playing with a doll."*
b) *"Peter was building figures with Lego blocks."*
c) *"John was enjoyed himself riding a bicycle."*

d) *"Paul told their children to come in because the*
   dinner was ready."

Summary text:

> *"The children were playing and their father told them*
> *to come in to have dinner."*

According to many studies, summarization is a difficult skill, because the gist of a text is often not present in the surface structure of the text (Friend, 2001) and the summarizing strategies which converts surface structure to the gist of a text are very complex. However, it is clear, that training improves students' skills in summary writing (Fan, 2010b; Karbalaei & Rajyashree, 2009; Pakzadian & Rasekh, 2013). Furthermore, it also helps students to pick out the main ideas from the text, and then express them in their own words. Hence, we carried out a study on students summaries in order to compare the students' performance in summary writing with the summarizing strategies that they used and we reported the study in the following section.

3.3 An Analysis on Students' Performance in Summary Writing

This section delineates the analysis of students' performance in summary writing. It aims to compare the students' performance in summary writing with the summarizing strategies that they used.

In this study, the following research questions are addressed:

1. Is there a correlation between summarizing strategies and students' performance in summary writing?

2. Does the number of summarizing strategies that the students used, affect the students' performance in summary writing?

3.3.1. Samples

In this section, we describe the data that is used in our study. All students' summaries (refer to Appendix C.1) and an article (refer to Appendix B.1) as a sample of the summary writing exercise were used in this study. Participants in this study were degree students (male, female) from Universiti of Teknologi Mara. The students' age is between 19 and 22.

3.3.2. Procedure

Two experts were identified to be subjects in this study. They are:

1. An English teacher with good reading skills and understanding ability in the English language as well as experience in teaching summary writing.
2. A lecturer with experience in using the skills in their teaching method.

The students were asked to read and summarize the given article. Then, each student worked alone and constructed an individual summary of the given article, which served as sample for this study.

Human expert assessment – two experts were asked to score the students' summaries on a scale of 0 to 100 with respect to how much the information in the original text overlap with the information in the student summary. Each of the two assessors independently produced scores for every summary. Finally, the student summaries were assigned the average of the scores that the human experts had allocated to each of the student's summary.

Inter-raters agreement – we used Pearson correlation coefficient as a measure of agreement between the two raters. The Pearson correlation coefficient for measuring the inter-raters agreement is 0.74. This value indicates that our assessors had good agreement for grading each of the student's summary.

Moreover, in order to compare the students' performance in summary writing with the summarizing rules that have been used by students in summary writing, the human experts were asked to identify summarizing rules from each student's summary. These summarizing strategies are determined as the following procedure:

1. Students' summaries were distributed between two human experts;

2. The human experts would deconstruct every student summary into a number of sentences, and then compare each sentence of student summary with all sentences from the original text to determine whether the sentences are semantically identical or not. Semantically identical sentences mean they share the same information or similar idea. The sentences from the original text that are semantically equivalent with the current sentence of student summary can be considered as the source sentences that have been contributed to produce the current summary sentence;

3. Given two sentences, the summary sentence and the source sentence, the experts determined the summarizing rules employed by students to produce the current sentence of summary text. To do so, they also utilized a guideline for identifying the summarizing rules.

The guidelines for identifying summarizing strategies are as follows:

i. **_Deletion_** – a summary sentence is created using deletion if the number of words in summary sentence is less than the number of words in the corresponding source sentence. The irrelevant and redundant information are removed.

ii. **_Paraphrase_** – a summary sentence is created using paraphrase if a word in the source sentence is replaced with a synonym word in the summary sentence.

iii. **_Sentence combination_** – summary sentence is created using sentence combination if two or more sentences from the source text are combined in the summary sentence.

iv. **_Generalization_** – a summary sentence is created using generalization if a list of similar items or actions from the source text are replaced using a general word or phrase in the summary text.

v. **_Topic Sentence Selection (TSS)_** – a summary sentence is created by TSS, if the sentence,

   1. *Title method;*

   2. *Location method;*

   3. *Key-words method;*

Tables 3.1 and 3.2 show some of the key words and stop words respectively.

Table 3.1: Examples of key words, extracted from the source text

Key words
"Water, Father, Shore, boat, Swim, Deep, Look, Day, Dive , fear, Terror, Try,"
"struggle, Faith, Grab, Ocean, Body, knew, Mouth, Reach, Push, Current, float",…

Table 3.2: Examples of stop words

Stop words
"About, above, across, after, afterwards, again, against, all, almost, alone,"
"already, also, although, always, am, among, amongst, amongst, amount, an,"
"another, any, anyhow, anyone, anything, anyway, anywhere, are, around, as,"
"along, and," ...

4. *Cue method.*

Table 3.3 shows some of these cue words that may appear in a sentence.

Table 3.3: Examples of Cue words

Cue words
"therefore, thus, consequently, hence, as a result, to conclude, in conclusion,"
"as a result, in short, to sum up, to summarize, to recapitulate, In consequence,"
"last of all, finally, to end, to complete, to bring to an end, to close"
"as a consequence of," ...

   vi.   ***Invention*** – a summary sentence is created using invention if one makes explicit topic sentences by using his or her own words to state the implicit main idea of the paragraphs.

vii.   ***Copy-verbatim*** – a summary sentence is created using copy-verbatim if all words in the summary sentence are found in the source sentence. The number of words in summary sentence is equal to the number of words in the source sentence.

The example of the summarizing rules, which were extracted from students' summaries and students' summaries scores are shown in Table 3.4.

Table 3.4: Summarizing strategies identified by the experts and students' summaries scores

Summaries	Summarizing strategies								Scores
	D	**S.C**	**G**	**P**	**TSS**	**Inv**	**CV**	**Total**	
1	7	2	0	0	8	0	1	18	45
2	7	1	0	0	10	0	4	22	52
3	13	2	0	1	13	0	1	30	54
4	11	1	0	0	9	0	2	23	44
5	15	1	0	0	15	0	0	31	53
Where,									
D: Deletion, **SC**: Sentence combination, **G**: Generalization, **P**: Paraphrase, **TSS**: Topic Sentence Selection, **Inv**: Invention, **CV**: Copy –verbatim.									

3.3.3. Results

The experimental data are subjected to a statistical analysis in order to explore the effectiveness of summarizing strategies such as deletion, sentence combination, topic sentence selection, paraphrase, invention and copy-verbatim, on the students' performance in summary writing. It also aims to provide answers to the two research questions of the study: 1) *Is there a correlation between summarizing strategies and students' performance in summary writing?* 2) *Does the number of summarizing strategies that the students used, affect the students' performance in summary writing?*

The study is conducted based on two hypotheses:

1. Null hypothesis (denoted H_0): There is no correlation between the summarizing strategies and students' performance in summary writing; and, the number of summarizing strategies does not affect the students' performance.

2. The alternative hypothesis (denoted H_A): There is a correlation between summarizing strategies and the students' performance in summary writing; and, the number of summarizing strategies affects the students' summarizing performance.

We used linear regression to analyse the correlation between the summarizing strategies and the students' performance. To analyse the correlation, Statistical Package software for Social Sciences (SPSS, version .17) was used.

Regression analysis is a statistical process for estimating the relationship between a dependent variable and an independent variable. Let *NSS* (independent variable) denote the number of summarizing strategies used by students in summary writing and *SP* (dependent variable) denote the students' performance in summary writing. More specifically, we want to understand how the typical value of the students' performance changes when the number of summarizing rules is varied. The relationship between *NSS* and *SP* is written as follows:

$$SP = \alpha\,(NSS) + \beta \tag{3.1}$$

Where,

The variable *SP* is named the "dependent" variable; *NSS* is named the "independent" variable; β is the "constant amount" and α is the "coefficient" of the variable *NSS*.

We performed a linear regression on the experimental data where the graph is presented in Figure 3.2. Each point (A) in the diagram represents an individual in the sample. This plot also shows the original data points along with the line providing the best fit through the points. The equation for the line is also given as follows:

$$SP = 1.3919 \, (NSS) + 11.382 \qquad (3.2)$$
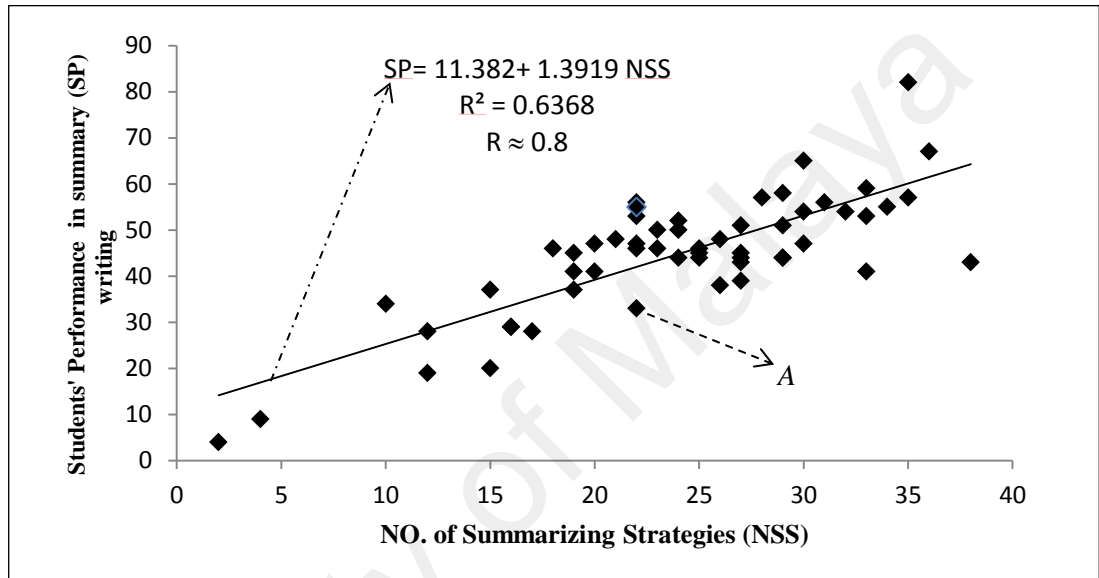
Figure 3.2: The correlation between summarizing strategies and the students' performance

Based on Figure 3.2 and the results in Table 3.5, we can conclude that there is a relationship between students' performance and the number of summarizing strategies. We explain each of them below.

- According to the categories of correlation values, as shown in Table 3.6, the linear correlation coefficient, $R \approx 0.8$ value shown on the graph indicates the goodness-of-fit for the line through the given points. This shows that there is a positive and strong correlation between students' summarizing performance and the number of summarizing strategies.

- Given the linear regression (Eq. 3.1), the coefficient of the *NSS* is bigger than zero ($1.392 > 0$). This means that there is a positive relationship between *NSS* and *SP*. In other words, when *NSS* increases, *SP* tends to also increase.

- Due to the results of Table 3.5, since $p - value = 0.000 < 0.05$, we shall reject the null hypothesis H_0 and accept the second hypothesis H_A.

Table 3.5: Statistical Results of Test for Regression -ANOVA on students' summaries

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	6205.443	1	6205.443	64.660	.000
Residual	3539.985	54	65.555		
Total	9745.429	55			

a. Predictors(Constant) NSS   b. Dependent Variable: SP

Table 3.6: Different types of correlation between two variables

Value (R)	Correlation
0	No correlation
0 to +1	Positive correlation
0 to -1	Negative correlation
(-, +) 0.5 or lower	(Negative, positive) Weak correlation
(-, +) 0.5 ... (-, +) 0.8	(Negative, positive) medium correlation
(-, +) 0.8 or higher	(Negative, positive) strong correlation

3.3.4. Discussion

Figure 3.2 and Table 3.5 display considerable results. The results proved that there is a strong relationship between students' performance in summary writing and the summarizing strategies they used. It shows that if the students know a number of summarizing rules, their skills in summary writing will be improved. These findings are in agreement with the previous studies of Lee (2010), Pérez, Bandera, León, and Cerván (2003) and Khoshsima and Nia (2014) which demonstrate that teaching summarizing strategies positively affects students' performance and writing quality. Thus, it can be suggested to teachers that summary writing should be taught to students. By teaching the proper steps, they learn how to use these summarizing rules. They also can understand what they are reading and how to select the main idea from the text.

As a conclusion, there is a strong relationship between the summary writing proficiency of students and the summarizing strategies used by the students. Their skills will be improved if they use proper summarizing strategies. Thus, the null hypothesis H_0 of this study is rejected by these findings. It seems that summary writing instruction based on the teaching summarizing strategies will help students to write quality summary.

3.4 Summary

This chapter outlines the steps to teach summary writing, the importance of teaching summary strategies, procedures of the study and finally the results of the study. We first discussed the importance of summary writing instruction, and then indicated the steps of training. Meanwhile, we compared the process of summarization with the basic summarization strategies. Finally, we carried out a study on students' summaries to examine the correlation between students' performance in summary writing and the summarizing strategies used by students. The present study contributes to the research

on the effects of collaborative text comprehension and summarizing strategies. It demonstrates that there is a strong relationship between summarizing strategies and students' performance. Thus, when the students are taught on how to use the summarizing rules, their summary writing skill will be enhanced.

CHAPTER 4

HEURISTIC RULES FOR IDENTIFYING SUMMARIZING STRATEGIES

4.1 Introduction

The task of training summarizing rules in support of producing an informative text summary is essential for students (Khoshsima & Nia, 2014; Lee, 2010; Pérez et al., 2003). [142]For this reason, several recent researches have focused on instruction of these rules. [142]Identifying summarizing strategies, used by student in summary writing, provides appropriate information to instructor such as students' ability to use summarizing strategies, how students use summarizing rules and their students' weakness in summarizing. As the process of identifying students' summarizing rules is very time consuming, a computer-assisted assessment will help teachers to conduct this task more effectively. Hence, the purpose of this chapter is to present the results of the analysis conducted on human−written summary in order to provide a set of rules for automatic identification of summarizing strategies used by students in summary writing.

4.2 Identifying summarizing strategies used in summary writing

To be more specific, we define the objective of this study as follows: Given a human−written summary sentences and original text sentences, the study will answer the following question:

- *How can the summarizing strategies be identified?*

For this purpose, an analysis has been done on human−written summary to collect a set of rules for identifying summarizing strategies that they used in producing a summary. This analysis is detailed as follows:

4.2.1 Samples

In this section, we describe the data that is used throughout our study. All students' summaries (refer to Appendix C.1) and an article (refer to Appendix B.1) as a sample of the summary writing exercise were collected from a school.

4.2.2 Procedure

Two experts are identified to be subjects in this study. The students were asked to read and summarize the given article. Then, each student worked alone and constructed an individual summary of the given article, which serves as sample for this study.

Human expert assessment – the human experts were asked to identify summarizing rules from each student's summary. These summarizing strategies are determined using the procedure as mentioned in section 3.3.2.

Table 4.1 shows the results of the analysis of the students' summaries. In particular, first column displayed the number of each sentence. The second column shows the SS, the third column presents the most relevant source sentences that were employed to generate a SS; and finally the last column displays the summarizing strategies that were used to generate each SS. . In appendix C.1, we show the results obtained from our study on the summary sentences.

This study aims to determine most relevant sentences from the original text for each summary sentence and identify the summarizing strategies used to construct the summary sentence. In appendix C.1, we show the results obtained from our study on the summary sentences.

Table 4.1: An analysis on summaries sentences

No. of sentence	Summary sentence	Original sentence	Summarizing strategy
1	"The currents kept pushing the boat further and further away."	"I took a couple of steps towards it, but the currents kept pushing the boat further and further away." (S_6)	• *Deletion*
2	"I plunged into the ocean and I knew I had overcome my fear."	"I plunged into the ocean and swam back to shore. (S_{33}) As my father proudly looked on, I knew I had overcome my fear." (S_{35})	• *Sentence combination*
3	"I dived and swam back to shore."	"I plunged into the ocean and swam back to shore." (S_{33})	• *Paraphrase*
4	"I was so traumatized."	"In the days that followed, I was so traumatized that I would not go near the water." (S_{25})	• *T.S.S (Beginning)* • *Deletion*
5	"He frantically searching for my body."	"He repeatedly dived under the water, frantically searching for my body." (S_{19})	• *T.S.S (End)* • *Deletion*
6	"I kicked hard, trying to remain above the surface."	"Panic-stricken, I paddled and kicked hard, trying to remain above the surface."(S_{12})	• *T.S.S (Title)* • *Deletion*
7	"My father was worried that the incident would scare me for life."	My father was worried that the incident would scare me for life." (S_{26})	• *Copy-verbatim*
8	"My father plunged and swam as hard as he could to the spot where I had gone under and frantically searching for my body."	"He dived in and swam as hard as he could to the spot where I had gone under. (S_{18}) He repeatedly dived under the water, frantically searching for my body." (S_{19})	• *Deletion* • *Sentence combination* • *T.S.S (End)* • *T.S.S (Title)* • *Paraphrase*

4.2.3 Results

This study aims to identify relevant sentences and the summarizing strategies employed to create the SS. Based on the analysis that we performed on the data, we obtained several conclusions to identify summarizing rules. Based on the previous literature (J. Allen, 1987; Pérez et al., 2005) we explain the following main observations.

1) Deletion strategy

The aim of deletion strategy is to "reduce without major loss"; whereby it eliminates unimportant phrases or words from a sentence, without losing the main idea of the original sentence. In fact, the deletion strategy eliminates a phrase or word from the original sentence if they are unnecessary and not related to the main topic. A phrase includes any of the sentence Stages such as a noun phrase, a verb phrase, a word and a prepositional phrase that can be deleted in the deletion rules. We can produce summary sentences by removing unnecessary phrases or words from the original sentences.

To identify the deletion strategy, we collected summary sentences and the corresponding sentences from the original text (Table 4.2 shows an example of these sentences that are extracted from Appendix C.1), then we found that some phrases or words have been removed from the original sentences and the number of word in summary sentences is less than the number of words in original sentence. Thus, a summary sentence is produced by deletion strategy rule if the length of the sentence is shorter than the length corresponding source sentence where some words or phrases have been removed from the original sentence.

Table 4.2: Deletion strategy

Summary sentence	Original sentence	Summarizing strategy
"He frantically searching for my body."	**"He** *repeatedly dived under the water*, **frantically searching for my body**."	• *Deletion*
"He took me to a jetty and challenged me to jump into the water."	*"When I was confident enough,* **he took me to a jetty and challenged me to jump into the water**."	• *Deletion*
"The currents kept pushing the boat further and further away."	*"I took a couple of steps towards it, but* **the currents kept pushing the boat further and further away."**	• *Deletion*

2) Sentence Combination Strategy

Sentence combination strategy merges phrases from two or more source sentences. Normally, the sentences that are combined by sentence combination strategy are mostly adjacent to each other in the source text. However, they can also be very far apart in the source text. The sentence combination strategy is widely employed by human summarizer (Jing, 2001). In addition, by merging the related sentences together as a single sentence, it produces informative details in summary text.

In this study, we also aim to identify how the sentence combination strategy can be identified. We analysed the example summary sentences that have been constructed by human summarizer using sentence combination strategy. After analysing the examples, we identified two main conditions that can be used to identify sentence combination strategy. First, sentence combination strategy employs two or more source sentences to create a single sentence in summary text. Secondly, the sentence combination strategy add conjunction words such as "and", "but", "or"," for", and "so" to combine two sentences into a single sentence. Table 4.3 shows an example of summary sentence constructed using sentence combination strategy.

Table 4.3: Sentence Combination strategy

Summary sentence	Original sentence	Summarizing strategy
"He took me to the edge of the water, *and* he taught me how to swim."	"One day, he took me to the edge of the water and urged me to go in with him.(S$_{27}$) In the following weeks, my father taught me how to swim."(S$_{29}$)	• *Sentence Combination*

3) Paraphrasing strategy

In paraphrasing strategy, a word from source sentence is replaced with a similar word or synonym word in the summary sentence. In other words, the summarizer uses other words to express the same meaning or information. Paraphrasing uses synonyms to change the words but not the meaning of the sentence. A sample of paraphrase strategy is shown in Table 4.4. In this sample, the word *grab* in the source sentence is replaced with a synonym word *catch*, which conveys the same meaning.

Table 4.4: Paraphrasing strategy

Summary sentence	Original sentence	Summarizing strategy
"I tried to *catch* the boat."	"In one huge step, I stretched my arm as far as it could go and tried to *grab* the boat."	• *Paraphrasing*

4) Topic Sentence Selection Strategy

To identify the important sentence, the following methods are considered:

    i.   Sentence location, containing the position of sentences within the source text. The sentence in the beginning or in the end of a paragraph is relevant to the main idea of the paragraph.

ii.   Relevant to title, including the sentence that one or more words of the title or major heading of a source text appear in the source sentence.

iii. Word frequency, it is the number of times a word occurs in the source text. Word frequency measures the importance of a word in the original text. The sentence containing one or more words with high frequency is more relevant and important.

iv.   Cue word, if a sentence includes cue words, such as "in summary" and "in conclusion".

A sample topic sentence selection strategy is shown in Table 4.5.

Table 4.5: Topic Sentence Selection (TSS) strategy

Summary sentence	Original sentence	Summarizing strategy
"I was so traumatised."	"In the days that followed, I was so traumatized that I would not go near the water."	• *TSS (beginning)*
"The currents kept pushing the boat further and further away."	"I took a couple of steps towards it, but the currents kept pushing the boat further and further away."	• *TSS (End)*
"I gasped for air in desperation, but only salty *water* filled my throat and nostrils." (S₁₅)	"I gasped for air in desperation but only salty *water* filled my throat and nostrils. "	▪ *T.S.S (Title word)*
"When I recovered, I opened my eyes to see my *father* crying." (S₂₃)	"He is crying."	▪ *T.S.S (Key word)*
"In one experiment, 84 low-income, minority high school juniors were randomly assigned to teachers who implemented direct instruction of summary skills or a control group of teachers did not provide explicit instruction on summarization strategies. *In conclusion*, the findings indicate that direct instruction of summarization skills has a significant effect on improving the skills of students."	"Findings indicate that direct instruction of summarization skills has a significant effect on improving the skills of students."	▪ *T.S.S (Cue word)*

5) Generalization strategy

Generalization strategy is to replace a general term for a list of words. This strategy includes two substitutions. *1)* The replacement of a general word for a list of similar items. *2)* The replacement of a general word for a list of similar actions.

A sample of generalization strategy is shown in Table 4.6.

Table 4.6: Generalization strategy

Summary sentence	Original sentence	Summarizing strategy
"I gasped for air, but only salty water filled them."	"I gasped for air in desperation, but only salty water filled my throat and nostrils."	• *Generalization*

6) Invention strategy

In invention strategy, a summarizer produces a topic sentence by using his or her own words to state the main idea of the paragraph. A sample of invention strategy is shown in Table 4.7.

Table 4.7: Invention strategy

Summary sentence	Original sentence	Summarizing strategy
"My father saved my life."	"He repeatedly dived under the water, frantically searching for my body."(S_{19}) "Then his hand brushed against my hair."(S_{20}) "He grabbed it and yanked my head out of the water."(S_{21}) "He pulled me to shore and started mouth-to-mouth resuscitation."(S_{22}) "When I recovered, I opened my eyes to see my father crying."(S_{23})	• *Invention*

7) Copy– verbatim strategy

In the copy – verbatim process, a summary sentence is created from the source sentence without any changes. This strategy is not part of the summarizing rules but

it is the strategy that is used by students. A sample of copy– verbatim strategy is shown in Table 4.8.

Table 4.8: Copy– verbatim strategy

Summary sentence	Original sentence	Summarizing strategy
My father was worried that the incident would scare me for life.	My father was worried that the incident would scare me for life. (S_{26})	• *Copy– verbatim*
I watched in horror as my boat floated out of reach.	I watched in horror as my boat floated out of reach. (S_7)	• *Copy– verbatim*

The final results are shown in the Table 4.9. The results show that around 90% of students use deletion rule, which removes some trivial parts of a sentence. It shows the deletion rule is a strategy frequently used by the students. This result is compatible with the previous research (Brown & Day, 1983). Brown and Day (1983) have conducted an experiment where they concluded that all various age groups used deletion strategy and it is not a difficult rule in producing summary sentences.

Although copy–verbatim is not a basic summarizing strategy, from this study, we also found that copy–verbatim was repeatedly used. Other common strategies that were used by more than 80% of the students were sentence combination and topic sentence selection strategy.

Paraphrasing is another strategy where students replace a word from a source sentence with a synonym. 39% of students used paraphrasing strategy. As shown in Table 4.9, only 7% of students used generalization strategy. Brown and Day (1983) from their experiments found that the young students used the generalization strategy less frequently, and when they tried to employ the generalization strategy

they mostly used it inefficiently. On the other hand, they found that experts used the generalization strategy perfectly.

The result also shows that less than 4% of the students used invention strategy. The invention strategy is employed when there is no explicit sentence in a paragraph, which represents the main idea of paragraph. In this case, a summarizer should produce a topic sentence by his or her own words to express the main idea of a paragraph. Thus, the invention strategy is the most complicated skill and the ability of using this strategy will be improved with age and experiences and it also develops along with growing older (Brown & Day, 1983). However, due to its difficulty, it is not taught in school. According to Brown and Day (1983), young students seldom employ the invention strategy, but experts utilize the invention strategy a lot more.

Table 4.9: Number of each summarizing strategy used by students

Summarizing Strategies	Frequency	Percentage
Deletion	55	98.21%
Sentence combination	51	91.07%
Paraphrase	22	39.28%
Topic sentence selection	55	94.38%
Generalization	4	7.14%
Invention	2	3.5%
Copy– verbatim	47	83.92%

A very important aspect of this study was how to determine the summarizing strategies employed by summarizer in summary writing. For this purpose, we considered the following features such as what the strategies are, where and when they can be used, and how they are employed. In other words, which summarizing strategies have been used and how these strategies applied to the source text for producing a summary

sentences. Using a set of heuristic rules, we are able to develop an algorithm that can be used to automatically identify summarizing strategies employed by summarizer to produce a summary. The obtained heuristic rules to identify each summarizing strategy is described in the next section.

4.3 Rules to Identify Summarizing Strategies

In this section, we are concerned to infer a set of rules for identifying automatically summarizing strategies that most commonly used by summarizer.

4.3.1 Deletion strategy

To identify the deletion strategy, we consider the following conditions:

4.3.1.1 *Sentence length*

Since the deletion strategy aims to omit unnecessary information from a sentence. Hence, the length of SS is always shorter than the corresponding original sentence.

Let S_s be a SS, O_s be an source sentence, let $\text{Len}(O_s)$ indicates the length of O_s and $\text{Len}(S_s)$ indicates the length of S_s. The following statement can be considered as a first rule:

$$Len\ (S_s)\ <\ Len\ (O_S) \tag{4.1}$$

The following rules were also considered in order to identify deletion strategy The following conditions are also considered to determine deletion strategy:

4.3.1.2 *Word overlapping*

Let $S_{summay} = \{W_1, W_2, \cdots W_N\}$ be a SS, where N is the number of $S_{summary}$, $S_{original} = \{W_1, W_2, \cdots W_M\}$ is a source sentence, where M is the number of words in

76

$S_{original}$. However, for each word from $S_{summary}$, the same word or the synonym word should be appeared in $S_{original}$. Hence, the following statement can be made:

$$\forall \; word \; \in S_{summay} \rightarrow \{same,\; synonym\}word \; \in S_{original} \quad (4.2)$$

4.3.1.3 *Syntactic composition* compares two sentences syntactically. It is used to determine whether the syntactic composition of two sentences is equal. For example, given two sentences:

$$\mathbf{S}_{original} = "\frac{He}{A} \text{ repeatedly dived under the water, frantically } \frac{searching}{B} \text{for my } \frac{body}{C} \text{ "}.$$

$$\mathbf{S}_{summay} = "\frac{He}{A} \text{ frantically } \frac{searching}{B} \text{ for my } \frac{body}{C} \text{ "}.$$

Suppose we select three words from sentence $S_{summary}$; A, B and C. If the word B occurred after A and the word C occurred after B, this composition should occur in sentence $S_{original}$. It means the word B must appear after word A and the word C must appear after word B in the $S_{original}$ sentence. Thus, the following statement can be made:

$$IF\; ((A,B \; and \; C \in S_{summay})\; \&\; A \rightarrow B \rightarrow C\;)\; THEN\; (A \rightarrow B \rightarrow C)\in S_{original}$$

$$(4.3)$$

We also consider the SSM as a rule to determine the deletion strategy. The SSM is measured based on the semantic and syntactic similarity measures. The reasons why we combine those similarity measures to compute the SSM and the details of the computing similarity between two sentences are explained in the next chapter. To sum up, the SSM are computed using the following processes (refer to chapter 5):

4.3.1.4 Similarity measure between two sentences

4.3.1.4.1 Given two sentences S_1 and S_2;

4.3.1.4.2 *Create the WS;*

4.3.1.4.3 *Compute the SSBS;*

4.3.1.4.4 *Compute the WOSBS;*

4.3.1.4.5 *Compute SSM.*

The following example is used to explain the computing similarity between two sentences.

1. Given two sentences S_1 and S_2.

    S_1="*Many dinosaurs, however, are middle-sized and about as big as a cow or horse.*"

    S_2="*dinosaurs are about as big as a cow.*"

2. Create the WS.

$$WS = \begin{cases} Many, dinosaurs, however, are, middle, sized, \\ and, about, as, big, as, a, cow, or, horse \end{cases}$$

3. Calculate SSBS using Eq. 4.4.

$$\vec{S}_1 = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$$

$$\vec{S}_2 = (0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0)$$

$$\text{Sim}_{\text{semantic}}(S_1, S_2) = \frac{\sum_{j=1}^{m}(w_{1j} \times w_{2j})}{\sqrt{\sum_{j=1}^{m} w_{1j}^2} \times \sqrt{\sum_{j=1}^{m} w_{2j}^2}}$$

$$= \frac{8}{\sqrt{8} \times \sqrt{15}} = \frac{8}{10.95} = 0.7305 < 1$$

4. Calculate the WOSBS using Eq.4.5.

$$\vec{O_1} = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15)$$

$$\vec{O_2} = (0, 1, 0 , 2 , 0, 0, 0, 3 , 4 , 5, 6, 7, 8 , 0 , 0)$$

$$Sim_{word\ order}(T, S) = 1 - \frac{||O_1 - O_2||}{||O_1 + O_2||} = 1 - \frac{85}{158}$$

$$= 1 - 0.5379 = 0.4621 \ < \ 1$$

5. Calculate SSM using Eq.4.6.

$$Sim_{\text{sentences}}(S_1, S_2) = \ 0.6926 < \ 1$$

In this study, we collected 163 SS_s generated by deletion strategy and the corresponding sentences source sentences. Then, we compute the SSM for each sentence pairs using equation (4.13). We found that the SSM for each sentence pairs in SC strategy is between 0 and 1, as displayed in Figure 4.1. Thus, the equation (4.4) can be considered as a rule to detect the deletion strategy:

$$0 < Similarity_{\text{sentences}}(S_1, S_2) < 1 \tag{4.4}$$

79
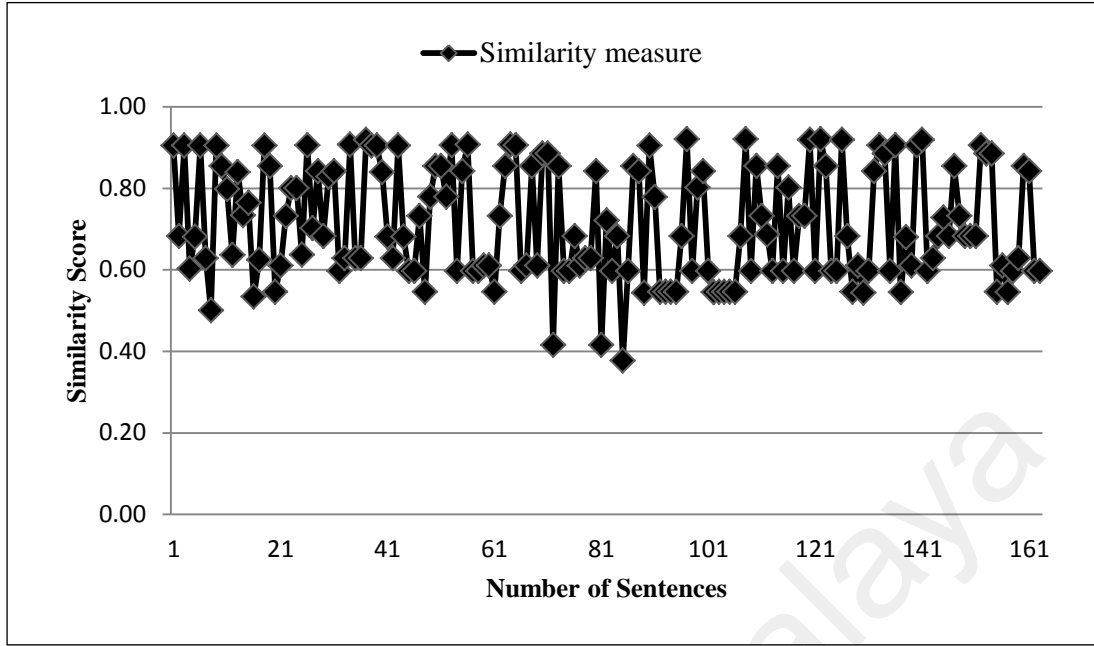
Figure 4.1: Sentence similarity measure in Deletion strategy

By analysing the SS_s created using deletion strategy, we found only one source sentence was used to create a SS. Hence, the equation (4.5) can be used to detect this strategy:

$$\text{Number of sentence } (N) = 1 \tag{4.5}$$

4.3.2 Topic Sentence Selection (TSS) Strategy

The TSS strategy aims to identify a sentence that represents the main idea of the paragraph. To determine TSS strategy, we consider 4 methods that are explained as follows.

4.3.2.1 Location method

In this study, we considered the use of position method to generate a SS. To do so, we analysed 560 SS. We found that significant sentence normally appears at the beginning as well as at the end of a paragraph. As a result, Figure 4.2 shows 49% of the important

sentences were appeared at the beginning of paragraphs and 51% of significant sentences were appeared at the end of paragraphs. These findings are in agreement with the previous studies of Fattah and Ren (2009) and Bawakid and Oussalah (2008).

The following process can be considered to the T.S.S strategy using location method:

i.   Create *SLL* using sentences that appeared at the beginning or at the end of a paragraph in source text.

ii. Identify the corresponding source sentence for each SS.

iii. Consider the equation (4.6) to determine T.S.S:

$$\text{IF } (S_{original} \in SLL) \quad \text{THEN } \{ Set\ TSS\ to\ 1;\ Otherwise\ Set\ to\ 0; \}$$

(4.6)
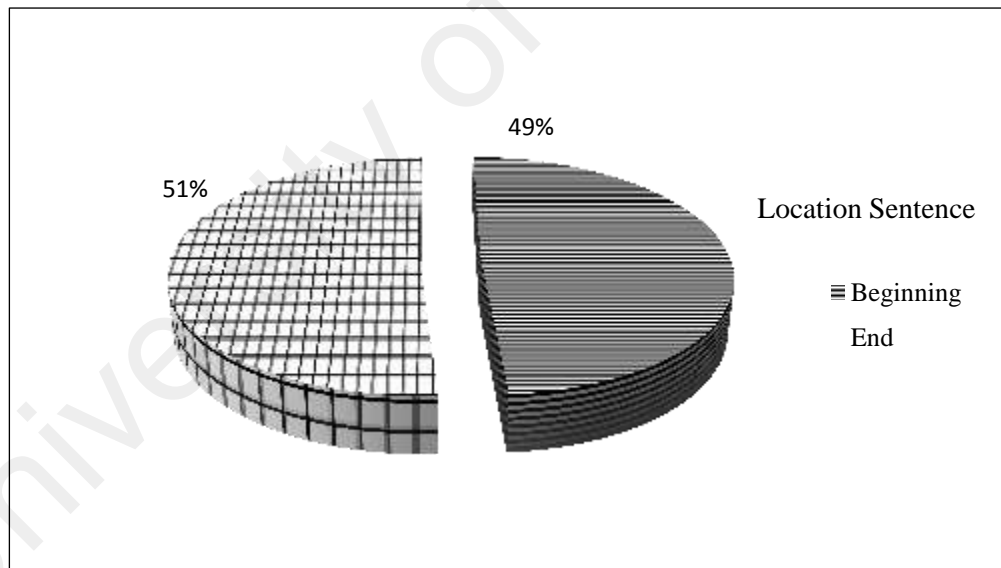
Figure 4.2: Use of Location Method amongst 56 summaries

4.3.2.2 Key word method

In this study, the Tf method is used to determine words with high frequency. Figure 4.3 displays words with high frequency.
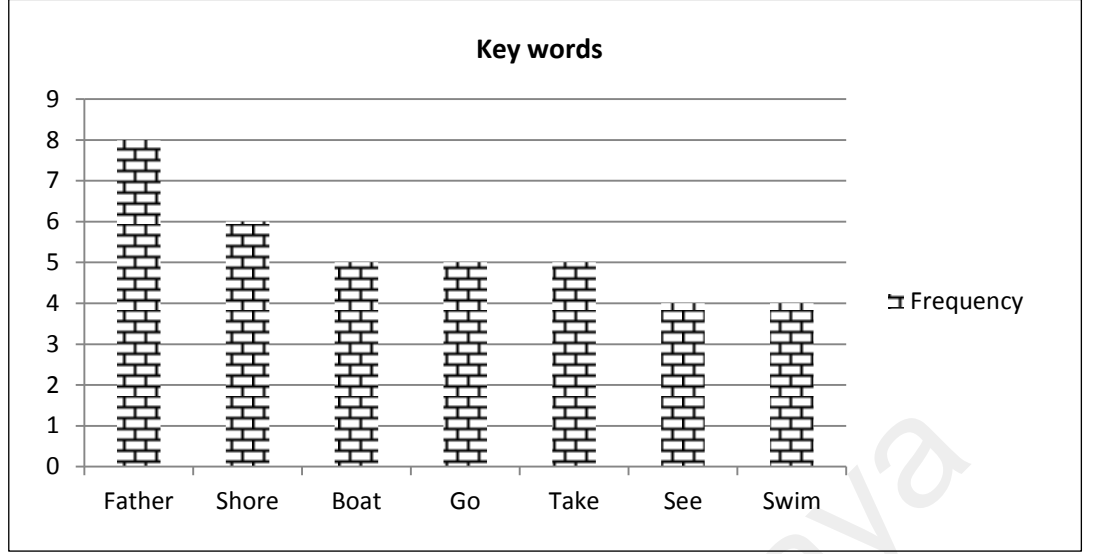
Figure 4.3: Frequency of keywords

In this study, we determine source sentences, which include title words, and were employed to create SS$_s$. Figure 4.4 presents the percentage use of each word, which have been employed to choose a significant sentence in T.S.S strategy.

The following process can be considered to the T.S.S strategy using keyword method:

     i.   Determine the frequency of each word (non-stop words)

    ii.   Create *KL* using top *N* words with high frequency

    iii.   Identify the corresponding source sentence for each SS

    iv.   Consider the equation (4.7) to determine T.S.S

$$\text{IF} \left(\exists \ word \ \in \ KL \ \ \text{and} \ \ word \ \in \ S_{\text{original}} \right)$$

$$\text{THEN} \ \{ Set \ TSS \ to \ 1; \ Otherwise \ Set \ to \ 0; \} \quad (4.7)$$

Where, S_{summay} is SS and S_{original} is a corresponding source sentence that was employed to generate the S_{summay}.
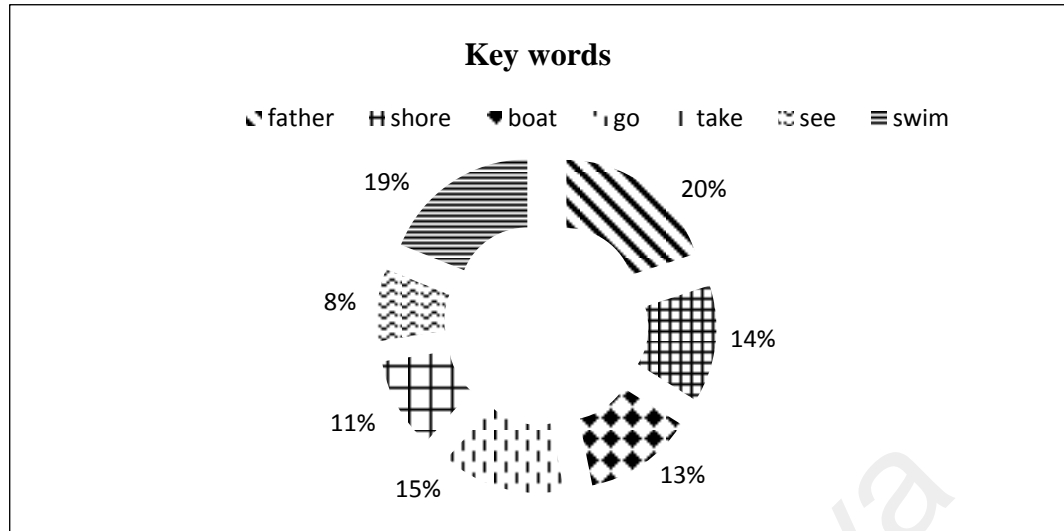
Figure 4.4: Use of keywords amongst 56 summaries

4.3.2.3 Title Method

In this study, we determine source sentences, which includes title words, and were employed to create SS_s. Figure 4.5 presents the percentage use of each title words, which have been employed to choose a significant sentence in T.S.S strategy.
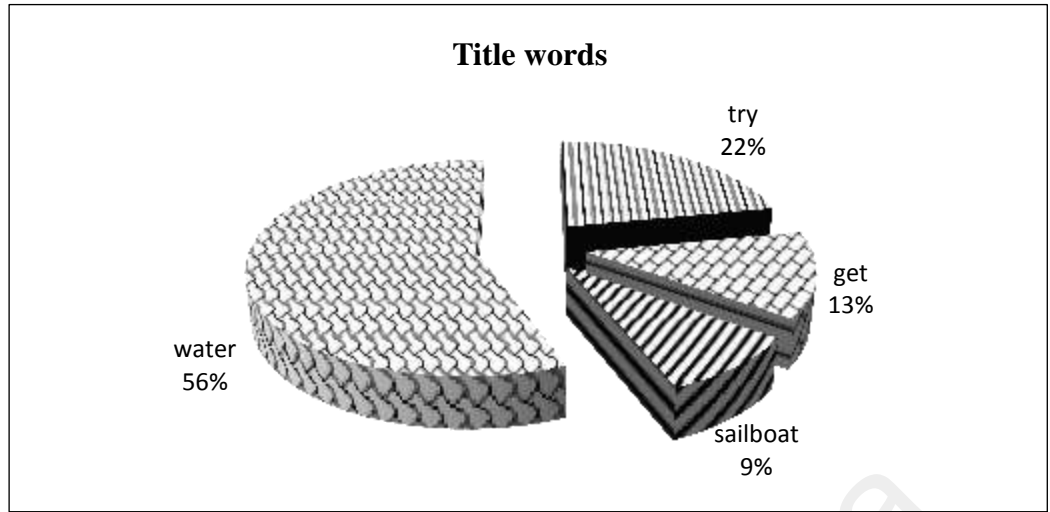
The following process can be considered to the T.S.S strategy using Title method:

     i.   Create *TL using title words* (non-stop words)

     i.   Identify the corresponding source sentence for each SS

     ii.   consider the equation (4.8) to determine T.S.S:

$$\text{IF} \left(\exists \ word \ \in \ TL \ \ \text{and} \ \ word \ \in \ S_{\text{original}} \right)$$

$$\text{THEN} \{ \ Set \ TSS \ to \ 1; \ Otherwise \ Set \ to \ 0; \} \quad (4.8)$$

Where, S_{summay} is SS and S_{original} is a corresponding source sentence that was employed to generate the S_{summay}.

83

Figure 4.5: Use of Title words amongst 56 summaries

4.3.2.4 "Cue method"

The following process can be considered to the TSS strategy using cue method:

    i.   Make *CWL* using the DM

    ii.   Identify the corresponding source sentence for each SS

    iii.   consider the equation (4.9) to determine TSS strategy

$$\text{IF}\left(\exists\ word \in CWL \quad \text{and} \quad word \in S_{\text{original}}\right)$$

$$\text{THEN}\{Set\ TSS\ to\ 1;\ Otherwise\ Set\ to\ 0;\} \qquad (4.9)$$

Where, S_{summay} is SS and S_{original} is a corresponding source sentence that was employed to generate the S_{summay}.
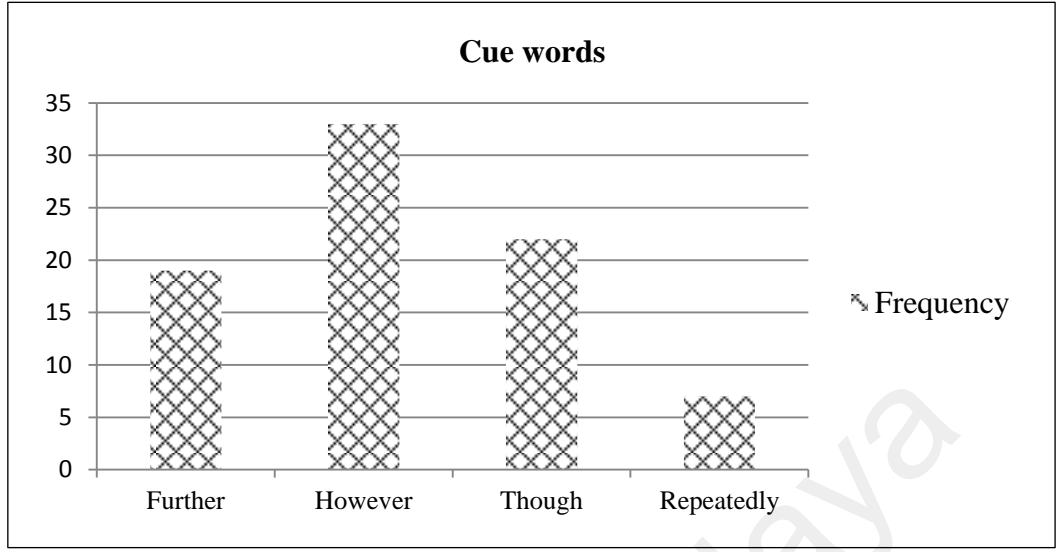
Figure 4.6: Frequency of cue words amongst 56 summaries

4.3.3 Paraphrasing strategy

We employed the following processes to determine the paraphrasing strategy:

1. Let $S_{summay} = \{W_1, W_2, \cdots W_N\}$ be a summary sentence and $S_{original} = \{W_1, W_2, \cdots W_M\}$ be a corresponding sentence of the source text that was employed to create the S_{summay}.

2. *AR* includes the root of each the root of each word of $S_{original}$.

3. *AS includes* the synonym of each word of $S_{original}$.

4. Retrieval *RW* For each word of S_{summay} using WordNet,

    i. If $RW \; \varepsilon \; AR$, then S_{summay} used paraphrase strategy; then jump to step 4; otherwise jump to step ii.

    ii. If $RW \notin AS$, then S_{summay} used paraphrase strategy; Stop; Otherwise continue by step iii;

    iii. Compute the SSBW (RW and words in $S_{original}$), (more details will be explained in chapter 5).

85

iv.   If exists a similar value, then S_{summay} used paraphrase strategy; Stop; Otherwise jump to 4;

4.3.4 Sentence Combination Strategy

In this study, we examined two features such as the number of source sentences combined in each SS and the SSM. For this purpose, we collected 105 SS created this strategy.

By analysing these sentences, we found that most SS_s are created using two or three sentences of the original text.

As shown in Figure 4.7, out of 105 SS produced using this strategy, 70 SS were created usually from two original sentences, 28 SS_s were created from 3 original sentences and 7 SS_s were created by 4 original sentences.

As a result from this study, the equation (4.10) can be considered as a rule to detect this strategy, SC:

$$N > 1 \qquad\qquad (4.10)$$

Where, N is the number of original sentences that were employed to create a SS.
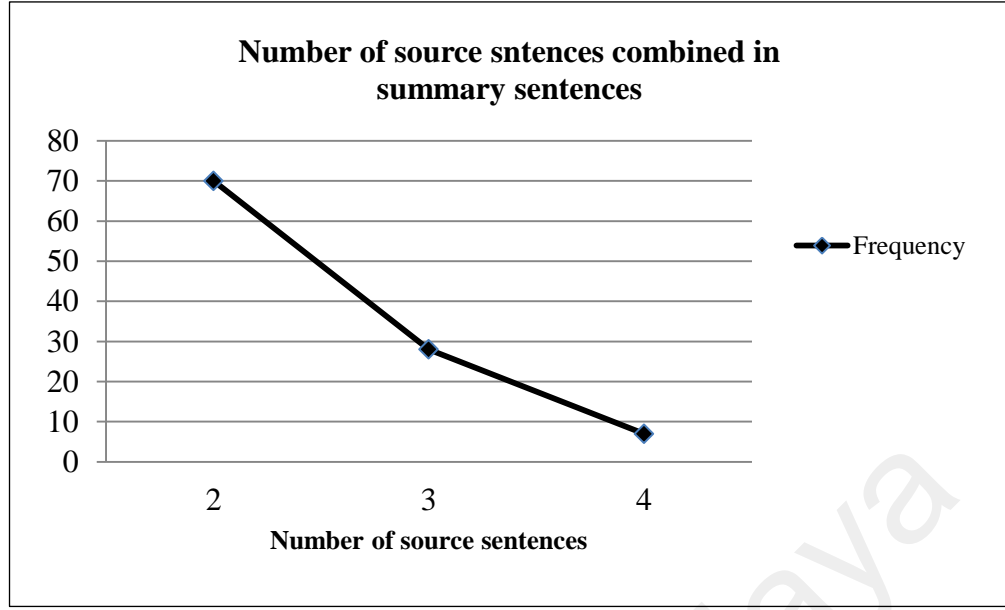
Figure 4.7: Number of source sentences combined in each summary sentence

We also consider the SSM as a rule to detect SC strategy. To compute SSM in SC strategy we perform the following steps. The details of the computing similarity are explained in the next chapter.

i. Given a SS $= \{P_1, P_2 \cdots P_N\}$, where P_1, P_2 and P_N are phrases from SS that came from T_1, T_2, and T_M respectively. T_1, T_2, and T_M are original sentences that were employed to generate the SS.

ii. Compute the SSM for each pair of sentences, such as $(T_1, SS), (T_2, SS) \cdots$, and (T_M, SS) through the following processes:

      i. Construct the WS;

      ii. Compute the SSBS;

      iii. Compute the WOSBS;

      iv. Compute SSM.

iii. Compute the average SSM using the equation (4.11):

$$Ave_{\text{similarity measure}} = \frac{\sum_{i=1}^{M} Sim(T_i, S_{summary})}{M} \tag{4.11}$$

Where, M is the number of source sentences.

The following example is used to explain how to compute SSM in SC strategy.

i. Given the following sentences:

$$SS = ``\frac{However, the\ water\ was\ just\ too\ deep}{P_1}\ and\ \frac{I\ felt\ myself\ sinking\ to\ the\ bottom}{p_2},$$

$$\frac{but\ only\ salty\ water\ filled\ my\ throat\ and\ nostrils}{P_3}\ ''.$$

$$T_1 = ``\frac{However, the\ water\ was\ just\ too\ deep}{P_1},\ the\ current\ was\ too\ strong\ and\ my$$

body was too weak to fight.''

$$T_2 = ``\ Terror\ overcame\ me\ as\ \frac{I\ felt\ myself\ sinking\ to\ the\ bottom}{P_2}\ ''.$$

$$T_3 = ``\ I\ gasped\ for\ air\ in\ desperation,$$

$$\frac{but\ only\ salty\ water\ filled\ my\ throat\ and\ nostrils}{P_3}\ ''.$$

ii. Calculate SSM for each pair sentences:

$$Sim_{sentences}(SS, T_1) = 0.3748$$

$$Sim_{sentences}(SS, T_2) = 0.3117$$

$$Sim_{sentences}(SS, T_3) = 0.2928$$

iii. Calculate the average similarity measure between sentences:

$$Ave_{similarity\ measure} = \frac{\sum_{i=1}^{M} Sim(T_i, SS)}{M}$$

$$= \frac{\sum_{i=1}^{3} Sim(T_i, SS)}{3}$$

$$= \frac{0.9793}{3}$$

$$= 0.3264$$

We collected 100 SS$_s$ generated by SC strategy and the corresponding source sentences. Then, we compute the SSM for each sentence pairs using equation (4.11). We found that the SSM for each sentence pairs in SC strategy is between 0 and 1, as displayed in Figure 4.8. Thus, the equation (4.12) can be considered as a rule to detect the current strategy:

$$0 \; < \; Ave_{\text{similarity measure}} = \; \frac{\sum_{i=1}^{M} Sim(T_i, S_{summary})}{M} \; < 1 \qquad (4.12)$$
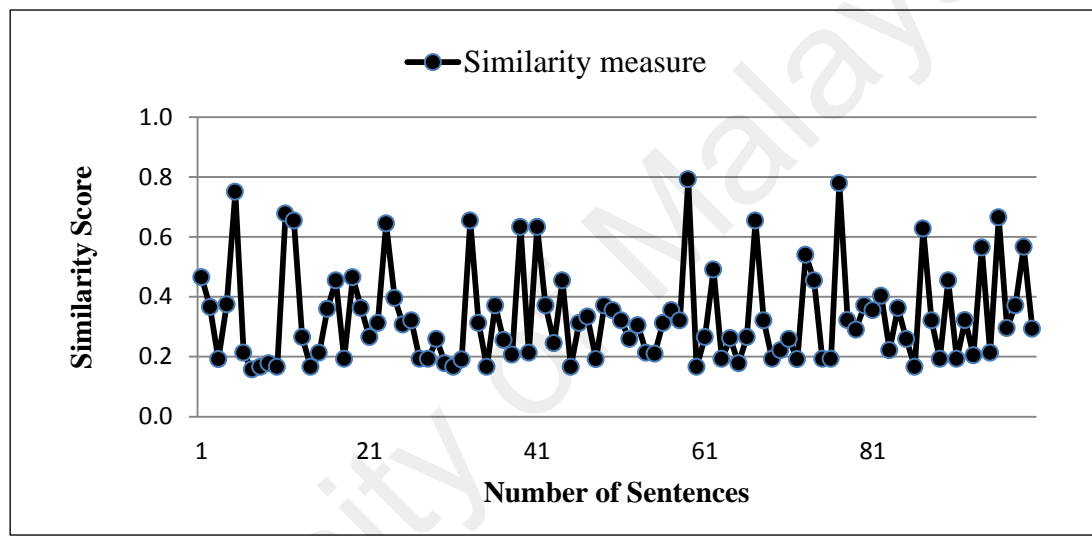
Figure 4.8: Sentence similarity measure in Sentence combination strategy

4.3.5 Copy– verbatim

By analysing the students' summaries, we found that most of the students used copy–verbatim to produce summary sentences.

The following rules are employed to detect the copy – verbatim strategy::

4.3.5.1 Sentence length

Since copy–verbatim strategy creates a SS using an original sentence without any changes; hence, the length of SS and a corresponding original sentence is always equal.

Let S_s be a SS, O_s be an source sentence, let $\text{Len}(O_s)$ indicates the length of O_s and $\text{Len}(S_s)$ indicates the length of S_s.

The equation (4.13) can be considered as a first rule:

$$Len \ (S_s) \ = \ Len \ (O_S) \tag{4.13}$$

4.3.5.2 "Similarity measure between sentences"

We also consider the SSM as a second rule to detect this strategy. the SSM are computed using the following processes (refer to chapter 5):

1. *Construct the WS;*
2. *Compute the SSBS;*
3. *Compute the WOSBS;*
4. *Compute SSM.*

We collected 80 SS$_s$ created by this strategy and the corresponding source sentences. Then, we computed the SSM for each sentence pairs. Finally, we found that the SSM for each sentence pairs in this strategy is equal to 1. Thus, the equation (4.14) can be considered as a rule to detect the current strategy:

$$Similarity_{\text{sentences}}(S_1, S_2) = 1 \tag{4.14}$$

4.3.5.3 "Total number of sentences"

By analysing the SS$_s$ created using this strategy, we found only one source sentence was used to create a SS. Hence, the equation (4.15) can be used to detect this strategy:

$$\text{Number of sentence } (N) = 1 \tag{4.15}$$

4.4 Summary

In this chapter, we analysed human summaries to determine a set of rules to in order to detect summarizing strategies automatically. We explored several rules to identify each summarizing strategies. In this study, the question centres on summarizing strategy such as "*How can the summarizing strategies be identified?*".

Our hypothesis is that information in both summary sentence and source sentence would help to identify summarizing strategies. Such information is used as a distinct feature for recognizing each summary strategy. We examined each feature using several summary and source sentences. The results of our investigation illustrate the effectiveness of such features for identifying each summary strategy. We also found that when features are combined, a more accurate rule will be produced for detecting each summary strategy.

CHAPTER 5

RELEVANCE DETECTION AND SUMMARIZING

STRATEGIES IDENTIFICATION ALGORITHM

5.1 Introduction

This chapter discussed the proposed algorithm in details. The aim of this algorithm is to determine the relevancy between source sentences and summary sentences, and also to identify the summarizing strategies automatically, which are used by summarizer to produce each summary sentence. The proposed algorithm is called **R**elevance **D**etection and **S**ummarizing **S**trategies **I**dentification **A**lgorithm (RDSSIA). The RDSSIA comprises three stages that constitute the backbone of the proposed algorithm which are "*pre–processing, intermediate– processing* and *post– processing*".

In RDSSIA, various steps are performed to identify text relevancy and summarizing strategies. The main steps are shown in Figure 5.1 and Figure 5.2 stated as follows:

1. Data Acquisition.

    This step can be considered as the algorithm inputs. In this step, a user is asked to upload an original text to be summarized. There is not limitation on the size of the text or the number of sentences included in the text.

2. Pre–processing: *Feature extraction.*

    This step aims to perform the basic linguistic analysis on both the original text and summary text. The *pre–processing* step consists of several processes which are *sentence segmentation, stop word removal, part of speech tagging, title word*

extraction, find location of sentences, stemming and tokenisation. We explain each of these processes in details in section 5.3.

3. Intermediate–processing: Sub Algorithm 1 (Sentences *Relevance Identification Algorithm).*

This step focuses to identify the source sentences, which were employed to create a SS. To do so, this step contains two main Stages which are sentence similarity computation Stage and sentence relevance detection Stage. The task of each Stage will be explained in section 5.4.

4. Post–processing: Sub Algorithm 2 (*Summarizing Strategies Identification Algorithm).*

This step identifies the summarizing strategies, which are used by summarizer. It includes the main Stage which is the detection of summarizing strategies. This Stage will be described in section 5.5.

In this chapter the following research questions are also addressed:

1. *How can the relevancy between summary sentences and the source sentence be detected?*

2. *How can RDSSIA to identify text relevancy and summarizing strategies summaries be formulated?*

3. *How does the RDSSIA work?*
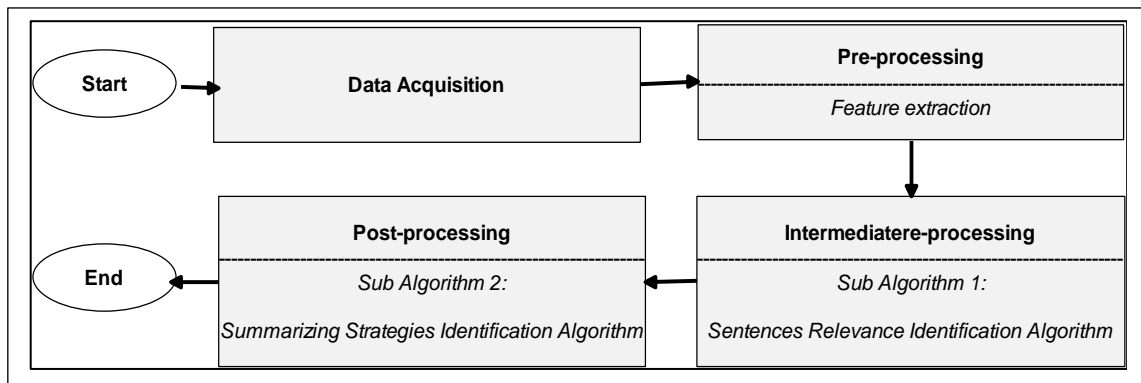
Figure 5.1: The proposed RDSSIA flow-diagram
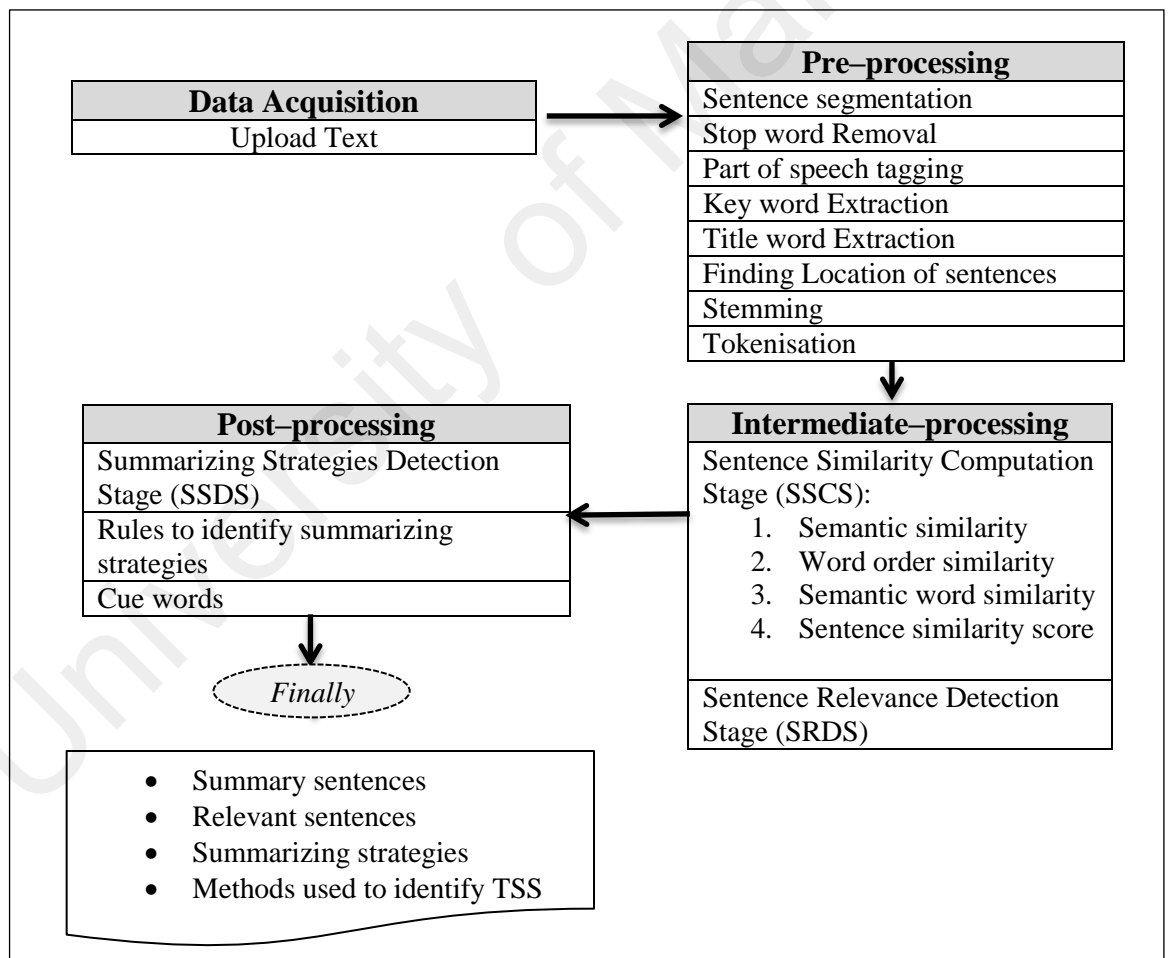
Figure 5.2: The proposed RDSSIA architecture

5.2 Development of the RDSSIA

The RDSSIA is a learning environment to assist teachers to identify summarizing strategies. It also aims to provide students with various feedbacks that can help them to improve their skills in summary writing and reading comprehension. The RDSSIA is aims to:

1. Identify whether a SS was produced using the source text.

2. Identify all relevant sentences from the source text that were employed to create SS.

3. Identify the summarizing strategies employed to create a SS.

The RDSSIA contains of two sub-algorithms, which are:

- *"Sentences Relevance Identification Algorithm"*

    This sub-algorithm includes several steps to identify original sentences that were used to create a SS. The corresponding process is shown in the "*intermediate– processing*" stage, as shown in Figure 5.3.

- *Summarizing Strategies Identification Algorithm*

    This sub-algorithm identifies the summarizing strategies that were employed to create a SS. The corresponding processes are presented in the *post–processing* stage, as shown in Figure 5.3.

Figure 5.3 shows the overview of the RDSSI algorithm, which includes: a) *Pre– processing*, b) *Intermediate–processing*, and, c) *Post–processing"*.

The explanation of each process of the core stages are explained in the next sections.
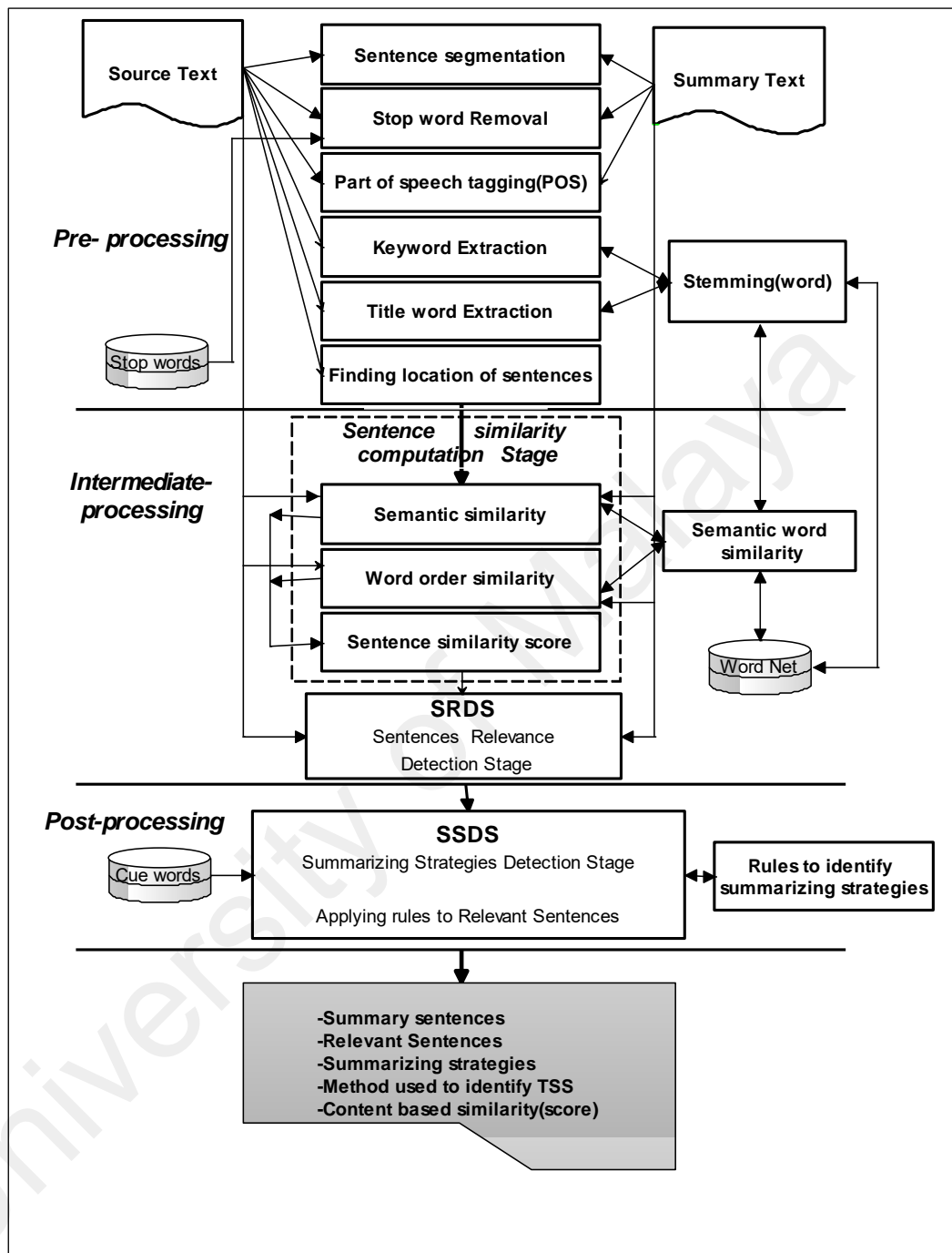
Figure 5.3: Overview of the development of the RDSSIA

5.3 "Pre–processing"

In this stage a basic linguistic analysis are applying on both the original text and summary text. Thus, it prepares both the original and summary text for further processing. This stage consists of a few processes which are:

- *Sentence segmentation* — in this process, the source text is split into sentences, which are the textual units considered for text relevance. A sentence ends with full stop (.) whereas a paragraph is ended by new line. Therefore, a paragraph consists of a group of sentences.

- *Tokenisation*— this process is used to identify each word in the source text, to calculate the frequency of each word and to determine stop words and nonstop words. In order to specify each word of source text, tokeniser is used.

- *Part-of-speech tagging*— in this process a part-of-speech tagger assigns to each word its morphological category such as (noun, verb, adjective, adverb, preposition, pronoun and conjunction). The result of this function is sent to sections, keyword extraction and title word extraction. We used an English part-of-speech tagger which was developed by Tsuruoka and Tsujii (2005) in University of Tokyo.

- *Stemming*— this process aims to identify words that have the same stem (e.g. *"go"* is the root of *"went"* and *"gone"*).

- *Stop word removal*— stop words includes words that occurred frequently in a text and are considered noisy terms, such as articles, propositions and conjunctions. Their removal can be helpful before the execution of a natural language processing task. Such removal is usually performed by word filtering with the aid of a list of stop words. To determine words with high frequency, this process is vital in our work. The list of English stop words (*564 words*) is

presented in Appendix D, which are extracted from the English stop word list (*http://jmlr.csail.mit.edu/papers/volume5/lewis04a/a11-smart-stop-list/english.stop* and *http://norm.al/2009/04/14/list-of-English-stop-words/#*).

- *Keyword extraction*— in this process, TF approach is used to determine words with high frequency.

- *Finding sentences location*— this process identifies the position of each sentence in an original text (first or last sentence of a paragraph or text).

- *Title word extraction*— in this process, all nouns and verbs are extracted from the title of a text.

WordNet is a lexical database for English which was developed at Princeton University (Miller & Charles, 1991). It includes 121,962 unique words, 99,642 synsets (each synset is a lexical concept represented by a set of synonymous words) and 173,941 senses of words. The lexical database, WordNet, is employed by the keyword extraction, the title word extraction functions and the sentence similarity computation Stage.

5.4 "Intermediate—processing"

This section is the core of the RDSSIA. In intermediate-processing all relevant sentences from original text for each SS are identified. To do so, the intermediate processing employed the SSCS and SRDS. We explain them as follows:

5.4.1. Sentence Similarity Computation Stage (SSCS)

The SSCS contains a computation model, SSCM, to compute the SSM. The (SSCM) is displayed in Figure 5.4. It presents the overall tasks to measure the SSM. The SSCM is responsible to determine all the relevant sentences from the source text for each SS.

This SSCM contains various stages; the process of each stage is explained in the next sections.
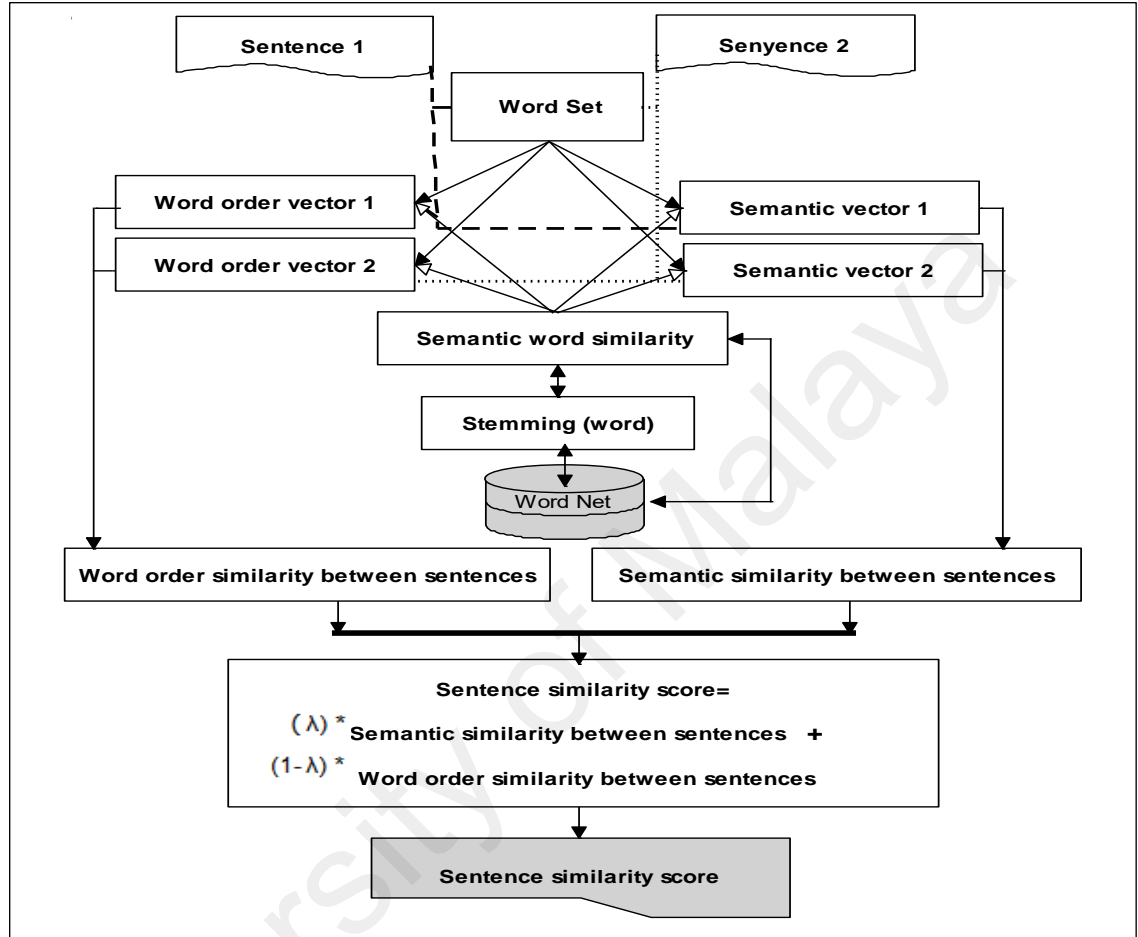
Figure 5.4: Sentence similarity computation model

5.4.1.1. Word Set

Given two sentences S_1 and S_2 , a "word set" is obtained using the following steps (Abdi, Idris, Alguliyev, & Aliguliyev, 2015, 2016):

1. "Two sentences are taken as input"

2. for each word, W, from S_1, the following steps are performed:

   i   Retrieval the root of each W using the WordNet.

   ii   If the $RW \ \varepsilon \ WS$ (word set), jump to step 2; otherwise, jump to step iii;

iii "If the $RW \notin WS$, then add RW to WS ; and then jump to step 2";

iv The same processes must be performed for Sentence 2.

The corresponding process is shown in algorithm 1.

Algorithm 1. The creation of "word set"

Input: Sentence 1, Sentence 2;

Output: $WS= \{W_1, W_2... W_n\}$, WS denotes an array that includes all distinct words from two sentences;

1: Let W be a word of the Sentence 1 or Sentence 2;

2: Let " RW be the root of word W, it is obtained using Word Net";

3: Let L be the length of Sentence1 or Sentence2;

4: Set $l = 0$;

5: For each W,

     i. l=l+1;

     ii. Get RW;

     iii. Look for RW in word set;

     iv. If "the RW was not in WS, then assign RW to WS";
      otherwise," jump to step 6";

6: Jump to step 5; iterate until $l \leq L$;

5.4.1.2. Semantic Similarity Between Words (SSBW)

Semantic word similarity (D. Lin, 1998; Tian, Li, Cai, & Zhao, 2010) is employed to produce a word-order and semantic-vector. Given two words, W_1 and W_2, the *SSBW* is calculated the using Eqs. (5.1) and (5.2).

The following equations (Abdi et al., 2016; Aytar, Shah, & Luo, 2008; D. Lin, 1998; Mihalcea, Corley, & Strapparava, 2006; Warin, 2004) are used to measure the *SSBW*:

$$IC\ (w)\ = 1 - \frac{Log(synset(w)+1)}{\log(\max\_w)} \tag{5.1}$$

$$Sim(w_1, w_2) = \begin{cases} \dfrac{2 \times IC\big(LCS(w_1, w_2)\big)}{IC(w_1) + IC(w_2)} & \text{if } w_1 \neq w_2 \\ 1 & \text{if } w_1 = w_2 \end{cases} \qquad (5.2)$$

Where *LCS* stands for the least common subsume. *LCS* of two words W_1 and W_2 is the set that contains all words of Synset (W_1) that also belong to Synset (W_2) (or equivalently, all words of Synset (W_2) that also belong to Synset (W_1). max_w is the number of words in Word Net, Synset (*W*) is the number of synonyms of word *W*, and *IC* (*w*) is the information content of word *W* based on the lexical database WordNet. Algorithm 2 presents the corresponding steps to calculate the semantic similarity measure between two words.

Algorithm 2. The Semantic similarity between words

Input: W_1, W_2, where W_1, W_2 are two words;

Output: similarity score between two words;

1: Let *LCS* denote least common subsume of two synsets;

2: Let RW_1 and RW_2 be the roots of words, W_1 and W_2 ;

 3: *If* $RW_1 = RW_2$, then two words are identical;

 4:Otherwise,

      i.   Assign the synset of W_1 to array1;

     ii.   Assign the synset of W_2 to array2;

    iii.   Assign the intersection of array1 and array2 to *LCS*;

    iv.   Assign the length of array1 to n_1;

     v.   Assign the length of array2 to n_2;

    vi.   Assign the length of the *LCS* to n_3;

   vii.   Compute the information content (*W_1*) using Eq. (5.1);

  viii.   Compute the information content (*W_2*) using Eq. (5.1);

    ix.   Compute the information content (LCS) using Eq. (5.1);

     x.   Compute the similarity score of W_1 and W_2 using Eq. (5.2);

5: Return the similarity score;

5.4.1.3. Semantic Similarity Between Sentences (SSBS)

The semantic—vector approach (Alguliev, Aliguliyev, & Mehdiyev, 2011; Aliguliyev, 2009; Li, McLean, Bandar, O'shea, & Crockett, 2006) is used to calculate the *SSBS*. The following steps are used to calculate *SSBS* (Abdi et al., 2015, 2016).

1. To create the semantic-vector.

    The semantic-vector for each corresponding sentence is created using word set. The length of the word set and the dimension of each vector are same.

2. To weight each cell of the semantic-vector.

    In semantic-vector each cell is weighted using the following steps:

    i   If a word, w, from word set exists in S_1, then the cell value is equal to 1. Otherwise, the next step will be performed;

    ii   If the w does not exist in S_1, then calculate *SSBW* between w and a words from S_1.

    iii If exists a similarity measure, then the cell value is equal to the highest similarity value.

    iv   If does not exist a similar value, then the cell value is equal to 0 (zero).

3. The semantic similarity measure is computed based on the two semantic-vectors, S_1 and S_2. The following equation is used to calculate the semantic similarity between sentences::

$$Sim_{semantic}\,(S_1\,,S_2) = \frac{\sum_{j=1}^{m}(w_{1j} \times w_{2j})}{\sqrt{\sum_{j=1}^{m}w_{1j}^2} \times \sqrt{\sum_{j=1}^{m}w_{2j}^2}} \qquad (5.3)$$

Where $S_1 = (w_{11}, w_{12}, \cdots, w_{1m})$ and $S_2 = (w_{21}, w_{22}, \cdots, w_{2m})$ are the semantic—vectors of sentences S_1 and S_2, respectively; w_{pj} is the weight of the jth word in vector S_p, m is the number of words.

Algorithm 3 presents the corresponding steps to create the lexical semantic vector.

Algorithm 3. Lexical Semantic Vector

Input: $\boldsymbol{S_1, S_2}$,"word set";

Output: semantic vector;

1: Let S be either senteence1 or sentence2;

2: Let W_t be a word of the word set;

3: "Let RW be the root of the word W_t, it is obtained using the Word Net";

4: Let W be a word of S;

5: Let SSM denotes the semantic similarity measure between words;

6: Let L be the length of S;

7: Set $l = 0$;

8: For each W_t ,

    i.   l= l+1;

    ii.   Get RW;

    iii.   Look for RW in S;

    iv.   If RW was in S, then set corresponding element in semantic vector to "1";

    v.   Otherwise,

        a. For each W,

           1. $SSM (W, W_t)$ is calculated using Eqs. (5.1) and (5.2);

           2. If $SSM \geq 0$ Then assign SSM to array1;

           3. Iterate until $l \leq L$;

        b. If array1=Null, then jump to step 9; otherwise,

        c. Select the most similarity value from array1;

        d. Set the corresponding element of the vector to the most value of similarity measure; set $l = 0$; jump to step8;

9: Assign '0" to the corresponding element of the vector; jump to step8; iterate until $l \leq L$;

5.4.1.4. Word Order Similarity Between Sentences (WOSBS)

Given two sentences S_1 and S_2, which include exactly the same words, but different word order. For example, two words (*father, child*) from S_1 appear in the reverse order in S_2.

S_1:" The <u>father</u> likes his <u>child</u>."

S_2: "The <u>child</u> likes his <u>father</u>."

Since these two sentences include the same words, the previous approach (*5.3.1.3 Semantic similarity between sentences*) can decide that S_1 and S_2 are equivalent. It is clear that the content of the two sentences is the same, but both sentences convey different meanings. This shows that, different word order leads to different meaning. Thus, it is necessary to consider the word order in the calculation for sentence similarity. Hence, to calculate the sentence similarity, we also used word order similarity measure.

We used syntactic-vector approach (Li et al., 2006) to measure calculate WOSBS. The following tasks were performed (Abdi et al., 2015):

1. Make the syntactic-vector for each sentences, S_1 and S_2;

   The syntactic-vector for each corresponding sentence is created using word set. The length of the word set and the dimension of each vector are same.

2. Weight each cell of the syntactic-vector.

   In syntactic-vector each cell is weighted using the index position of the words in the corresponding sentence. The weight of cells is calculated using the following steps:

i   If a word, w, from word set exists in S_1, then the cell value is equal to the index position of the current word in S_1. Otherwise, the next step will be performed;

ii   If the w does not exist in S_1, then calculate *SSBW* between w and a words from S_1.

iii   If exists a similarity measure, then the cell value is equal to the index position of the word in S_1 with the high similarity value.

iv   If does not exist a similar value, then the cell value is equal to 0 (zero).

3. The syntactic similarity measure is then computed based on two syntactic-vectors, S_1 and S_2, and the similarity measure is computed using the following equation:

$$Sim_{\text{word order}}(S_1, S_2) = 1 - \frac{||O_1 - O_2||}{||O_1 + O_2||} \tag{5.4}$$

Where $O_1 = (d_{11}, d_{12}, \cdots, d_{1m})$ and $O_2 = (d_{21}, d_{22}, \cdots, d_{2m})$ are the syntactic vectors of sentences S_1 and S_2, respectively; d_{pj} is the weight of the jth cell in vector O_p.

Algorithm 4 presents the corresponding steps to create the word-order vector.

Algorithm 4. Word order Vector

Input: sentence1, sentence2,"word set";

Output: Lexical vector;

1: Let S be either senteence 1 or sentence 2;

2: Let W_t be a word of the word set;

3: "Let RW be the root of the word W_t, it is obtained using the Word Net;"

4: Let W be a word of S;

5: Let SSM denotes the semantic similarity measure;

6: Let L be the length of S;

7: Set $l = 0$;

8: For each W_t,

      i. $l = l+1$;

     ii.   Get RW;

     iii.   Look for RW in S;

     iv.   If RW was in Sen, then set corresponding element in vector to index position of
           word in S;

     v.   Otherwise,

          a.   For each W,

              1. $SSM (W, W_t)$ is calculated using Eqs. (5.1) and (5.2);

              2. If $SSM \geq 0$ *Then assign SSM to array1;*

              3. *Iterate until $l \leq L$;*

          b.   If array1=Null, then jump to step 9; otherwise,

          c.   Select the most similarity score from array1;

          d.   Set the corresponding element of vector to index position of word with the
             most similarity score; set $l = 0$; jump to step8;

    9: Assign "**0**" to the corresponding element of the vector; jump to step 8; iterate until $l \leq L$;

5.4.1.5. Sentence Similarity Measurement (SSM)

In this section, we describe the hybrid similarity measure. By "hybrid", we refer to the

combinations two different similarity measures: semantic similarity and syntactic

similarity. This is based on the notion that the semantic and syntactic construction

contributes to the understanding of meaning of a sentence (Achananuparp, Hu, Zhou, &

Zhang, 2008; Kanejiya, Kumar, & Prasad, 2003; Li, McLean, Bandar, O'shea, &

Crockett, 2006; Lin, Ng, & Kan, 2011; Pérez et al., 2005; Wiemer-Hastings & Wiemer, 2000; Wiemer-Hastings & Zipitria, 2001). Thus the overall sentence similarity is defined as a linear combination (Eq. 5.5) of semantic similarity and syntactic similarity.

$$\text{Sim}_{\text{sentences}}(S_1, S_2) = \lambda \cdot \text{sim}_{\text{semantic}}(S_1, S_2) + (1 - \lambda) \cdot \text{sim}_{\text{wordorder}}(S_1, S_2)$$

(5.5)

Where $0 < \lambda < 1$ is the weighting parameter (also called parameter balancing, damping factor, regularization parameters, combination parameter, weighting parameter, trade-off parameter, adjusted parameter, weight coefficients, tuning factor, hybrid parameter and valued parameter), Specifying the relative contributions to the overall similarity measure from the semantic and syntactic similarity measures. In other words, λ is the hybrid parameter in the interval (0, 1), to make effective use of both semantic similarity and syntactic similarity.

Since syntax plays a subordinate role for semantic processing of text (Li et al., 2006; Wiemer-Hastings & Wiemer, 2000) and a sentence similarity measure performs the best when semantic measure is weighted more than syntactic measure, usually $\lambda > 1 - \lambda$ in (Eq. 5.5) and the high value of λ is given to similarity measure. The value of λ in the (Eq. 5.5) is set empirically. For this purpose, to find the optimal value for λ we perform a grid search through all values from 0.0 to 1.0 using step sizes of 0.1. Our intent is to optimize λ value for the best performance. Experimentally it is proved that $\lambda = 0.8$ gives best results (refer to section 6.3.1.2).

- It is clear that with bigger value of λ, more importance is given to the semantic similarity compared to syntactic similarity.
- On the other hand, with smaller magnitude of λ, the syntactic similarity will be more importance.

107

- If λ=0.5 the semantic and syntactic similarity measures are assumed to be equally important.

In fact, if the user tends the semantic similarity to be the main concern, λ should be higher. Besides, if syntactic similarity is the main concern, the user is encouraged to set (1- λ) higher value than λ. Finally, if the user tends the both syntactic similarity and semantic similarity to be equally important, the λ and (1- λ) should be equal (λ=0.5).

5.4.2. Sentences Relevance Detection Stage (SRDS)

The main task of this Stage is to identify original sentences that were employed to create a *SS*.

Let $T_{Original\ Text} = \{S_1, S_2 \cdots S_N\}$ represent all source text, N is the number of sentences. S_s shows a *SS*. Then, let

$Arr_{Relations} =$

$\{(S_1, S_s, Value_{sim(S_1,S_s)}), (S_2, S_s, Value_{sim(S_2,S_s)}) \cdots (S_M, S_s, Value_{sim(S_M,S_s)})\}$ represent source sentences that are associated with S_s, $M \leq N$. We aim to identify a set of the sentences from $Arr_{Relations}$ that were employed to create S_s. $Arr_{Relevant\ sentences}$ includes a subset of the sentences $Arr_{Relations}$. Algorithm 5 presents the process to identify these sentences. The algorithm mainly includes the following steps:

Algorithm 5. Selection source text sentences

Input: array₁: includes all source sentences that have relationship with a *SS*;

Output: array₂: contains sentences that were employed to produce a *SS*;

1: Let *S* be a *SS*;

2: Let *T* be a sentence of array₁;

3: Let *MSS* denotes a relation with most similarity score between two

    sentences *S* and *T;*

4: Select *MSS* ;

5: Remove all common words from *S* and *T;*

6: If length of   *S* is equal to 'Null', then

          i.   Assign   *T* to array₂;

          ii.   Remove *T* from array₁; Exit;

7: Otherwise,

          i.   Call Algorithm 2 *(S and T with remaining words);*

          ii.   Remove all similar words between two sentences (*S, T*);

8:If length of *S* is equal to 'Null', Then

          i.   Assign   *T to array₂;*

          ii.   Remove *T* from array₁; Exit;

9: Otherwise,

          i.   Generate the new sentence S'' using remaining words of the

              sentence S' ;

          ii.   Assign *T* to array2;

          iii.   Remove *T* from array1;

          iv.   Generate the new $\boldsymbol{array'_1}$ using remaining cells of the array₁ ;

          v. S'' and $\boldsymbol{array'_1}$ are sent to *SSCS*

Step 1. RDSSIA choses a relationship from $Arr_{Relations}$ with the high similarity value, ($Value_{sim(S_1,S_s)}$). Let S_1 is a sentence of $Arr_{Relations}$ has relation to S_s. Thus, these two sentences for more process are taken into next step.

Step 2. In this step, first the common words from sentences S_1 and S_s are eliminated; then, if $h(S_s) = 0$, it means only sentenceS_1 was employed to produce S_s, and S_1 is

considered as a relevant system. If $length\ (S_s) \neq 0$, then the algorithm proceeds to the next step.

Step 3. Let Let S_1' and S_s' represent S_1 and S_s with their remaining words, respectively.

The SSBW of words of S_s' and S_1' is computed using Eqs. (4.1) and (4.2). If the SSBW exists, the similar words would be eliminated. If $length\ (S_s') = 0$, it means only sentenceS_1 was employed to produce S_s, and S_1 is considered as a relevant system.

If $length\ (S_s') \neq 0$, it means the S_s was produced using two or more original sentences. Therefore, the S_1 is considered as a relevant sentence and the algorithm proceeds to the next step.

Step 4. In current step, to detect other source sentences that were employed to produce S_s, the $S_s^{''}$ with the remaining words of S_s' is sent to the *SSCS*.

5.5 Post—processing

The main task of this step is to identify summarizing strategies. This step efforts to answer the following questions:

1. *Which summarizing strategies have been employed to produce a summary sentence?*

2. *How can a T.S.S strategy be determined?*

3. *What are the methods employed to determine T.S.S strategy?*

The overall processes for applying a set of heuristic rules to determine the summarizing strategies are explained as follows:

5.5.1. Identifying Summarizing Strategies Used in Summary Writing

5.5.1.1. Deletion, Sentence combination, Copy-verbatim Strategies

Given two texts, summary and source text, let $S_s = \{W_1, W_2 \cdots W_K\}$ be a SS and $Arr_{RS} = \{(T_1, S_s, P_1), (T_2, S_s, P_2) \cdots (T_N, S_s, P_M)\}$ includes all the source sentences that were employed to generate the S_s, k is the number of words in S_s, M is the number of phrases in the S_s, T_N is the Nth sentence from the source text and (T_N, S_s, P_M) shows that the Mth phrase of sentence S_s comes from the Nth sentence from the source text. The process to determine aforementioned strategies are as follows:

Step 1. If $N = 1$, then the RDSSIA tries to determine the deletion and copy-verbatim strategy using *step 2*, otherwise, it tries to determine the S.C combination strategy using *step 3*.

Step 2. The RDSSIA checks the following statements to identify copy-verbatim and deletion strategy.

$$State_{CV} = \Big((N = 1) \wedge \big(Len(T) = Len(S_s)\big) \wedge (Sim\,(T, S_s) = 1)\Big) \qquad (5.6)$$

$$State_{Del} = \Big((N = 1) \wedge ("Len(T) > Len(S_s)) \wedge (0 < Sim\,(T, S_s") < 1)\Big) \qquad (5.7)$$

Where T represents a sentence of Arr_{RS} and $Sim\,(T, S_s)$ shows the SSM between T and S_s. $Len(T)$ and $Len(S_s)$ show the length of sentence T and S_s respectively.

The S_s employed the copy-verbatim strategy (Eq. 5.6), if $N = 1$, the $Len(T) = Len(S_s)$, and the $Sim\,(T, S_s) = 1$.

The S_s employed the deletion strategy (Eq. 5.7), if $N = 1$, the $Len(T) > Len(S_s)$, and the $Sim(T, S_s) < 1$. In other hands, the algorithm considers the *syntactic composition* and *word overlapping* in both sentences S_s and T (refer to section 4.3.1).

$$\forall \ word \ \in S_s \rightarrow \{same, \ synonym\} \ word \ \in T \tag{5.8}$$

IF ((W_1, W_2 and $W_3 \in S_s$) & $W_1 \rightarrow W_2 \rightarrow W_3$) THEN ($W_1 \rightarrow W_2 \rightarrow W_3$) $\in T$ (5.9)

Step 3. if $N > 1$, it means two or more than two sentences were used to produce S_s. However, if $N > 1$ and $Sim(T, S_s) < 1$ (Eq. 5.10), then the S_s employed the S.C strategy:

$$State_{sentence \ combination} = \left((N > 1) \wedge \left(0 < \frac{\sum_{j=1}^{i} Sim(s_{rj}, S_s)}{TN} < 1 \right) \right) \tag{5.10}$$

Since the S_s includes phrases from two or more sentences from source text; hence, each phrase of S_s can be considered separately for identifying other summarizing strategies.

5.5.1.2. Paraphrasing Strategy

Given two sentences, let $S_{summay} = \{W_1, W_2, \cdots W_N\}$ be a *SS*, *N* is equal to the number of words in the $S_{summary}$, $S_{RS} = \{W_1, W_2, \cdots W_M\}$ a sentence of $Arr_{Relevant \ sentences}$ that is employed to produce the $S_{summary}$, where *M* is equal to the number of words in the S_{RS}.

$A_{Root} = \{W_{R1}, W_{R2}, \cdots W_{RN}\}$ contains the root of each word of $S_{summary}$, where W_{Rj} is the root of jth word in $S_{summary}$.

$B_{Synonym} = \{W_1, W_2, \cdots W_K\}$ contains the synonym of each word of the $S_{summary}$. First, the algorithm determines the root and the synonyms for each word of S_{RS} using WordNet, then assign them to A_{Root} and $B_{Synonym}$, respectively.

Second, the algorithm obtain the root of the word (RW) using the WordNet. If $\varepsilon\ A_{Root}$, then the algorithm tries for next word; otherwise, it looks for RW in $B_{Synonym}$. If the search result is true, it means $S_{summary}$ employed the paraphrase strategy.

5.5.1.3. Topic Sentence Selection Strategy: Cue, Title, Keyword, Location methods

Given two sentences, S_{summay} and S_{RS}. Let,

S_{summay} be a SS and S_{RS} be a sentence of $\text{Arr}_{\text{Relevant sentences}}$ that is employed to create the S_{summay};

- $L_{cue\ word} = \{CW_1, CW_2, \cdots CW_N\}$ denotes a list of cue words;

- $L_{key\ word} = \{KW_1, KW_2, \cdots KW_k\}$ denotes a list of keywords;

- $L_{title\ word} = \{TW_1, TW_2, \cdots TW_M\}$ denotes a list of title words;

- $L_{sentence\ location} = \{(S_1, L_B, L_E), (S_2, L_B, L_E), \cdots (S_j, L_B, L_E)\}$ denotes the position of the sentences in the original text, where L_B and L_E shows the first last sentence of a paragraph, respectively. (S_j, L_B, L_E) displays that the jth sentence, S, from original text can be the first or the last sentence of a paragraph.

The process to identify the T.S.S strategy using various methods is explained as follows:

- Title method

  The S_{summay} employed title method, if the S_{RS} includes a word of $L_{title\ word}$.

- "Key-word method"

The S_{summay} employed key-word method, if the S_{RS} includes a word of $L_{key\ word}$.

- Location method

The S_{summay} employed location method, if the S_{RS} includes a word of $L_{sentence\ location}$.

- "Cue method"

The S_{summay} employed cue method, if the S_{RS} includes a word of $L_{cue\ word}$.

The final results of the algorithm are the summary sentences, the relevant sentences, the summarizing strategies employed to create the SS and the method employed to identify T.S.S strategy.

5.6 How RDSSIA Works?

This section aims to show how the algorithm works by using a simple example. We have a summary sentence, S_1, which has relation with sentences T_1, T_2 and T_3 from the source text. To find which original sentence and what summarizing strategies were used to create S_1, we consider the following cases.

i.   Case 1: A summary sentence is produced using only one sentence from the source text.

ii. Case 2: A summary sentence is produced using more than one sentence from the source text.

The description of each case is as follows:

5.6.1. Case 1: A summary sentence is created using only one sentence from the source text

Supposed the sentences, T_1, T_2, T_3 and S_1 are as follows:

S_1 : *"The currents kept pushing the boat further and further away."*

T_1: *"I took a couple of steps towards it, but the currents kept pushing the boat further and further away."*

T_2: *"However, the water was just too deep, the current was too strong and my body was too weak to fight."*

T_3: *"When I recovered, I opened my eyes to see my father crying."*

The following process are used to identify which original sentence and what summarizing strategies have been used to create the *SS*, S_1.

1) Stop Words removal

In this step the stop words such as prepositions, conjunctions are removed from sentences. Table 5.1 shows the sentences after stop words removal process:

Table 5.1: Sentences after stop word removal

Sentences	
S_1	*"Currents kept pushing boat"*
T_1	*"took couple steps currents kept pushing boat "*
T_2	*"water deep current strong body weak fight"*
T_3	*"Recovered opened eyes see father crying"*

2) Similarity measure calculation

  i. Create a "word set*"*

As discussed in the previous sub section (5.3.1), creating a word set is the first step to calculate the similarity measure. To do so, consider two sentences S_1 and T_1. According to the algorithm 1, the first step is to assign all words of sentence S_1 to an array or the word set. Then by a loop, each word of sentence T_1 is added to the array if the word was not in the array. Table 5.2 shows an example of this step when applied to sentences S_1 and T_1.

Table 5.2: Example of a word set

Sentences	Word Set (WS)
$S_1=$ *"Currents kept pushing boat"*	*steps, currents, pushing boat, kept, took, couple*
$T_1=$ *"took couple steps currents kept pushing boat"*	

  ii. Calculate semantic similarity between two sentences

Using algorithm 3, we can create the lexical semantic-vector (D_1 *and* D_2) for both sentences (S_1 and T_1). According to the algorithm, to create lexical semantic— vector D_1 for sentence S_1, we utilize the "word set" and sentence (*S1: "Currents kept pushing boat further further away"*). To do that, for each term (e.g. $w=$*"steps"*) of "word set" we locate it in sentence S_1. If the word appears in sentence S_1, we set the weight of corresponding word '$W=$*steps*' in vector D_1 to "1". Otherwise, we should compute similarity measure between current word and all words from sentence S_1. Then, the weight of corresponding word in vector D_1 is set to highest similarity value. When a word from "word set" doesn't exist in sentence S_1 and also, doesn't have any similarity

value between the current word and all words from sentence S_1, the weight of the corresponding word in vector D_1 is set to "0".

The semantic─vector D_2 for sentence T_1 is created according to the aforementioned steps. We calculated semantic similarity between sentences (S_1 and T_1) using semantic vectors (D_1 and D_2) and equation (5.3), as shown in Figure 5.5.

$D_{S1} = (0, 1, 1, 1, 1, 0, 0)$

$D_{T1} = (1, 1, 1, 1, 1, 1, 1)$

$\text{Sim}(S_1, T_1) = \frac{(0\times1+1\times1+1\times1+1\times1+1\times1+0\times1+0\times1)}{(1+1+1+1+1+1+1)\times(1+1+1+1)} = \frac{4}{(\sqrt{7})\times(\sqrt{4})} = 0.7559$

Figure 5.5: Semantic similarity calculation

Table 5.3 shows the semantic similarity measure between any pair of the sentences.

Table 5.3: Semantic similarity measure between sentences

Sentence pair	Word Set	Semantic similarity score
$S_{1\ \&\ T_1}$	*Currents, kept, pushing, boat, Took, couple, steps, kept, pushing, boat*	0.7559
$S_{1\ \&\ T_2}$	*Currents, kept, pushing, boat, Water, deep, strong, body, weak, fight*	0.1889
$S_{1\ \&\ T_3}$	*Currents, kept, pushing, boat, Recovered, opened, eyes, see, father, crying*	0

The Algorithm 4 presents some steps to make a word order—vector. However, for each word W from the "word set", if the word exists in sentence S_1, the weight of word W is equal to the index position of the corresponding word from S_1. Otherwise, it should find the most similar word in sentence S_1 using SSBW. The weight of the word, W, is set to the index position of the most similar word from sentence S_1. In some cases, if there is no similar or equivalent word in sentence S_1, the weight of the word W sets to "0".

The word order—vectors (O_1 and O_2) for sentence (S_1 and T_1) is created according to the aforementioned steps. Then, the word order similarity measure between sentences (S_1 and T_1) is calculated using semantic vectors (O_1 and O_2) and equation (5.4). Figure 5.6 shows how we calculate WOSBS (S_1 and T_1).

WS={*steps, currents, pushing, boat, kept, took, couple*}

S_1 = Currents$_{[1]}$ kept$_{[2]}$ pushing$_{[3]}$ boat$_{[4]}$

T_1 = Took$_{[1]}$ couple$_{[2]}$ steps$_{[3]}$ currents$_{[4]}$ kept$_{[5]}$ pushing$_{[6]}$ boat$_{[7]}$

O_{S1}= (0, 1, 3, 4, 2, 0, 0)
O_{T1}= (3, 4, 6, 7, 5, 1, 2)

$Sim_{wordorder}(S_1, T_1) = 1 - \frac{18}{38} = 0.5263$

Figure 5.6: Word order similarity calculation

Table 5.4 shows the word order similarity score between any pair of sentences.

Table 5.4: Semantic similarity score between sentences

Sentence pair	Index position of words	Word order similarity score
$S_{1\ \&\ }T_{1}$	S1 = Currents$_{[1]}$ kept$_{[2]}$ pushing$_{[3]}$ boat$_{[4]}$ T1 = Took$_{[1]}$ couple$_{[2]}$ steps$_{[3]}$ currents$_{[4]}$ kept$_{[5]}$ pushing$_{[6]}$ boat$_{[7]}$	0.5263
$S_{1\ \&\ }T_{2}$	S1 = Currents$_{[1]}$ kept$_{[2]}$ pushing$_{[3]}$ boat$_{[4]}$ T1 = Water$_{[1]}$ deep$_{[2]}$ current$_{[3]}$ strong$_{[4]}$ body$_{[5]}$ weak$_{[6]}$ fight$_{[7]}$	0.0526
$S_{1\ \&\ }T_{3}$	S1 = Currents$_{[1]}$ kept$_{[2]}$ pushing$_{[3]}$ boat$_{[4]}$ T1 = Recovered$_{[1]}$ opened$_{[2]}$ eyes$_{[3]}$ see$_{[4]}$ father$_{[5]}$ crying$_{[6]}$	0

iv.   Calculate SSM

We also calculated SSM using equation (5.5). Table 5.5 shows the results.

Table 5.5: Similarity measure between sentences

Sentence pair	Similarity score
$S_{1\ \&\ }T_{1}$	0.7099
$S_{1\ \&\ }T_{2}$	0.1616
$S_{1\ \&\ }T_{3}$	0

3) *Sentences Relevance Detection Stage (SRDS)*

This step aims to identify which original sentences have been employed used to produce the *SS*. The process to identify these source sentences are presented in Algorithm 5. According to the algorithm, a relation with the most similarity score is chosen. As shown in Table 5.5, the pair of sentences which have the most similarity score is S_1 *and* T_1. Thus, this pair of sentences is selected.

Table 5.6: Example of identifying relevant sentence

Sentence
S_1 = ~~Currents~~ ~~kept~~ ~~pushing~~ ~~boat~~
T_1 = Took couple steps ~~currents~~ ~~kept~~ ~~pushing~~ ~~boat~~

As shown in Table 5.6, all words, which have been shared by two sentences, will be removed; then, If the length (S_1) = 0, it shows that S_1 contains a phrase from original sentence, and sentence T_1 is employed to produce the S_1.

If the length (S_1) ≠ 0, the similarity measure between the remaining words from the two sentences (S_1 *and* T_1) is calculated; then all similar words between the two sentences are removed. Finally, If the length (S_1) = 0, it shows that T_1 from source text was employed to produce sentence S_1 from the summary text. If the length (S_1) ≠ 0, it indicates that S_1 includes phrases from two or more source sentences.

We will explain this case in the next sub section (5.6.2). Since sentence T_1 was used to create sentence S_1, sentence T_1 is taken to the next process to identify the summarizing

strategies employed to create S_1. On the other hand, since only one sentence has been used to create sentence S_1 the value of $N = 1$.

4) Summarizing strategies identification

According to the results in Figure 5.7, the sentence T_1 *was* used to produce S_1, the value of N is equal to 1, and the similarity measure between two sentences S_1 and T_1 is 0.7559. In this section, the following steps are taken to determine summarizing strategies.

$S_1 =$ "Currents kept pushing boat"

$T_1 =$ "Took couple steps currents kept pushing boat"

$N= 1$, (number of sentences)

$Sim (S_1, T_1) =0.7559$

$Len(T_1) = 7$

$Len(S_1) = 4$

Figure 5.7: Summary information from sentences S_1 and T_1

- **Deletion**

According to the equation (5.7), the sentence $S1$ used deletion strategy.

Equation (5.7):

$$State_{Del} = \left((N = 1) \wedge (Len(T) > Len(S)) \wedge (0 < Sim\,(T,S) < 1)\right)$$

$$= \left((1) \wedge (7 > 4) \wedge (0 < 0.7559 < 1)\right)$$

According to the equations (5.8) and (5.9) each word of SS can be found in original sentence, and the SS and the corresponding original sentences have same syntactic composition.

- **Copy-verbatim**

  According to the equation (5.6), the sentence S_1 did not use copy-verbatim strategy.

  Equation (5.6):

  $$State_{CP} = \left((N = 1) \wedge \left(Len(T) = Len(S_s)\right) \wedge (0 < Sim\,(T, S_s) = 1)\right)$$

  $$= \left((1) \wedge (7 = 4) \wedge (0 < (0.7559) = 1)\right)$$

- **Sentence combination strategy**

  According to the equation (5.8), the sentence S_1 did not use sentence combination strategy.

  Equation (5.8):

  $$State_{sentence\ combination} = \left((N > 1) \wedge \left(0 < \frac{\sum_{j=1}^{i} Sim\left(s_{r_j}, S_s\right)}{TN} < 1\right)\right)$$

  $$= \left((1 > 1) \wedge (0 < (0.7559) < 1)\right)$$

- **Paraphrasing strategy**

  Since any word from sentence T_1 was not replaced with a similar word or synonym in sentence S_1, therefore the sentence S_1 did not use paraphrasing strategy.

- **TSS strategy**

In topic sentence selection, several methods can be employed to choose significant information from the original text. The process to detect each method is explained as follows:

i.   Title method

Supposed the words, as shown in Figure 5.8, are title words. The sentence S_1 used title word method if sentence T_1 contains at least one of the title words.

> *Try, get, sailboat, water*

Figure 5.8: Samples of Title words

ii.   Key word method

Figure 5.9 shows the words with high frequency. Sentence S_1 employed key word method if sentence T_1 contains at least one the key words.

> *"Water, Father, Shore, boat, Swim, Deep, Look,   Dive , fear, Terror, Try, struggle, Faith, Grab, Ocean, Body, knew, Mouth, Reach, Push, Current, float"*

Figure 5.9: Examples of key words, extracted from the source text

Since sentence T_1 includes keywords (*push and boat*), the sentence S_1 used keyword method.

iii.   Cue method

Several cue words are displayed in Figure 5.10. If sentence T_1 contains at least one of the cue words, it is concluded that sentence S_1 used cue method. According to the content of sentence T_1, sentence S_1 did not use cue method.

> *"therefore, thus, consequently, hence, as a result, to conclude, in conclusion, to close, last of all, finally, to end, to complete, to bring to an end, to sum up, to summarize, to recapitulate, In consequence,…"*

Figure 5.10: Examples of Cue words

iv.   Location method

Sentence S_1 used location method if the sentence T_1 appeared at the beginning or at the last sentence of a paragraph. Since sentence T_1 is the last sentence of a paragraph in a text (refer to Appendix B.1), the sentence S_1 used location method.

5.6.2. Case 2: A summary sentence is created using more than one sentence from the source text

In some cases, a *SS* is produced using more than one sentence. Supposed we have the following sentences.

S_1 *: "He took me to the edge of the water and he taught me how to swim."*

T_1*: "In the following weeks, my father taught me how to swim."*

T_2*: "One day, he took me to the edge of the water and urged me to go in with him."*

T_3*: "When I was confident enough, he took me to a jetty and challenged me to jump into the water."*

124

To determine which source sentence and what summarizing strategies were employed to create *SS*, S_1, we applied following steps:

1) Stop Words removal

After the stop words removal process, the sentences will be as follows:

S_1 *: "Took edge water taught swim"*

T_1*: "Following weeks father taught swim"*

T_2*: "Day took edge water urged go"*

T_3*:" Confident took jetty challenged jump water"*

2) Similarity measure calculation

Table 5.7 shows the similarity measure between each pair of sentences.

Table 5.7: Similarity measure between sentences

Sentence pair	Similarity score
$S_{1 \& } T_1$	0.4400
$S_{1 \& } T_2$	0.5048
$S_{1 \& } T_3$	0.3365

3) Sentences Relevance Detection Stage (SRDS)

According to the Algorithm 5, a relation with the most similarity score is chosen. As shown in Table 5.7, the pair of sentences which have the most similarity score is *S_1 and T_2.*Thus, this pair of sentences is selected.

125

Firstly, all similar words from the two sentences, S_1 and T_2 will be removed.

$$S_1 : \text{~~Took edge water~~ taught swim}$$

$$T_2: \text{Day ~~took edge water~~ urged go}$$

Thus, the remaining words of two sentences are as follows. Let S_1' represent sentence S_1 and sentence T_2' represent sentence T_2.

$$S_1': \text{taught swim}$$

$$T_2': \text{Day urged go}$$

Then, if the length (S_1') is not equal zero, the SSBW of sentences S_1' and T_2' would be calculated. Algorithm 2 presents the steps for measuring semantic similarity between two words. This algorithm receives two words (W_1 and W_2) as the input and returns a similarity score. According to the steps of Algorithm 2, we need to determine the synonym of two input words and to find LCS from two words. We also need to determine the number of synonyms of each word (W_1 and W_2) and least common subsume. Finally, the similarity score is computed by equations (5.1) and (5.2). The steps to calculate the similarity measure between words are as follows.

   i.   Retrieval synonyms of each word using WordNet

Synset (*taught)* = {*teach, learn, educate, train, instruct*}

Synset (*swim*) = {*swim, swimming, float*}

Synset (*urged*) = {*urge, impulse, urge on, press, exhort, recommend, cheer, inspire*}

Synset (*day*) = {*day, daytime, daylight*}

Synset (go) = {go, leave, come, proceed, travel, cross, pass, get on, grow, expire, walk, gait, ambulate, stride, traipse, depart, become, proceed, move, run, get going}

ii. Determine Number of Synonyms of each Word (NSW)

NSW $_{taught}$ =5; NSW $_{swim}$ =3; NSW $_{urged}$ =8; NSW $_{day}$ =3; NSW $_{go}$ =21;

iii. Determine LCS of two words and its length

LCS (taught, urged) =Null; *LCS (taught, day)* =Null; *LCS (taught, go)* =Null;

LCS (swim, urged) =Null; *LCS (swim, day)* =Null; *LCS (swim, go)* =Null;

iv. Calculate similarity score

Example:

To calculate the similarity, *SIM*, between word *taught* and *day:*

$$IC\ (taught) = 1 - \frac{Log(synset(w)+1)}{\log(max_w)}$$

$$= 1 - \frac{\log(5+1)}{\log(122000)}$$

$$= 0.8470$$

$$IC\ (day) = 1 - \frac{Log(synset(w)+1)}{\log(max_w)}$$

$$= 1 - \frac{\log(3+1)}{\log(122000)}$$

$$= 0.8816$$

Since the Least Common Subsume (LCS) from two words (*taught* and *day*) is equal to zero, the result of following equation is also equal to zero.

$$Sim(taught, day) = \frac{2 \times IC(LCS(w_1, w_2))}{IC(w_1) + IC(w_2)}$$

$$= \frac{2 \times (0)}{0.8470 + 0.8816}$$

$$= 0$$

Using aforementioned steps, we also calculate the similarity measure between other words. Finally, the result shows that there is no similarity between the words of S_1' and T_2'.

SIM (taught, urged) =;   *SIM (taught, day) =0*;   *SIM (taught, go) =0*;

SIM (swim, urged) =0;   *SIM (swim, day) =0*;   *SIM (swim, go) =0*;

As a result, this case shows more than one sentence was employed to produce the *SS, S₁*. To find other sentences that were used to create *S₁, the similarity measure between* sentences S_1', *T₁* and *T₃*, as shown in in Figure 5.11 is calculated. The sentence *T₂* is recorded as a sentence has been employed to produce *S₁*.

S$_1'$ *: taught swim*

T_1: "*Following weeks father taught swim*"

T_3: "*Confident took jetty challenged jump water*"

Figure 5.11: Examples of sentences S$_1'$, T$_1$ and T$_3$

The similarity measure between sentences, (S_1', T_1) and (S_1', T_3) is calculated using The *Sentence Similarity Computation Stage*. Then the similarity scores between sentences are sent to the *Sentences Relevance Detection Stage*. Table 5.8 present the results.

Table 5.8: Similarity measure between sentences

Sentence pair	Similarity score
S_1' & T_1	0.5725
S_1' & T_3	0

As shown in Table 5.8, the pair of sentences which have the most similarity score is S_1' and T_1. Given the two sentences S_1' and T_1, the similar words are eliminated from the two sentences.

S_1' : ~~taught swim~~

T_1: *Following weeks father* ~~taught swim~~

Thus, the remaining words of two sentences are as follows,

S_1' : Null

T_1: *Following weeks father*

As we can see, the length of sentence S_1' is equal to zero. Hence, sentence T_1 was also used to produce sentence S_1. Finally, it was found that original sentences (T_1 and T_2) were employed to produce the *SS, S₁*. On the other hand, since more than one sentence had been used to create sentence S_1, the value of *N* (*Number of sentences*) was bigger than 1.

4) Summarizing strategies identification

Using the results, as shown in Figure 5.12, this section aims to identify the summarizing strategies used to produce *S₁* as follows.

$$S_1 : \frac{Took\ edge\ water}{P_1} \frac{taught\ swim}{P_2}$$

T_1: *Following weeks father* $\frac{taught\ swim}{P_2}$

T_2: *Day* $\frac{took\ edge\ water}{P_1}$ *urged go*

N >1, (number of sentences)

$Ave\ _{Sim\ (S1,T1)} = 0.5062$

Figure 5.12: Summary information from sentences S_1 and T_1

- **Deletion**

As shown in Figure 5.12, the sentence S_1 includes two phrases, where phrase P_1 originates from sentence T_2 and phrase P_2 originates from sentence T_1. Hence, the following sentences and phrases are considered in order to identify deletion strategy.

P_1:" took edge water"

T_2: "Day took edge water urged go"

N= 1, (number of sentences)

Sim (P_1, T_2) =0.6545

$Len(T_2) = 6$

$Len(P_1) = 3$

P_2: "taught swim"

T_1: *"Following weeks father taught swim"*

N= 1, (number of sentences)

Sim (P_1, T_2) =0.5725

$Len(T_1) = 6$

$Len(P_2) = 2$

According to the following statement (equation (5.7)), the sentence S_1 is produced using deletion strategy.

$$State_{Del} = \left((N = 1) \wedge (Len(T) > Len(P)) \wedge (0 < Sim\,(T,P) < 1)\right)$$

According to the equations (5.8) and (5.9) each word of summary phrases P_1 and P_2 is found in source sentence T_2 and T_1 respectively. The phrases P_1 and P_2 and the corresponding original sentences T_2 and T_1 have same syntactic composition.

- **Copy-verbatim**

  As shown in Figure 5.12, we obtained the following results:

  $Len(T_2) > Len(P_1)$ and *Sim (P_1, T_2) <1*

  $Len(T_1) > Len(P_2)$ and *Sim (P_2, T_1) <1*

  Thus, according to the following statement (equation (5.6)), the copy-verbatim strategy is not used to produce sentence *S1*.

$$State_{CP} = \left((N = 1) \wedge (Len(T) = Len(P)) \wedge (0 < Sim\,(T,P) = 1)\right)$$

- **Sentence combination strategy**

  As shown in Figure 5.12, the average of the similarity measure is less than 1 and the number of sentences is bigger than 1. Thus, according to the following

131

statement (equation (5.8)) the sentence S_1 is produced using the sentence combination strategy.

$$State_{sentence\ combination} = \left((N > 1) \wedge \left(0 < \frac{\sum_{j=1}^{i} Sim\left(S_{r_j},\ S_s\right)}{TN} < 1 \right) \right)$$

- **Paraphrasing strategy**

  Since any word from sentences T_1 and T_2 was not replaced with any similar word or synonym in phrases P_2 and P_1, respectively, the paraphrase strategy was not used to produce sentence S_1.

- **Topic sentence selection strategy (TSS): cue, title, keyword, location methods**

  i. Title method

     As shown in Figure 5.12 sentences T_1 and T_2 were used to produce sentence S_1. To determine title method, we consider whether the title words, as shown in Figure 5.8, appeared in sentences T_1 or T_2. As a result, the sentence T_2 contains a title word, *'water'*, therefore, the title method and *TSS* strategy were used to produce sentence S_1.

  ii. Keyword method

     The sentence S_1 is created by keyword method if each of sentences T_1 or T_2 includes one or more of the key words (refer to Figure 5.9). Since the sentence T_1 and T_2 include keywords (*try* and *swim*), hence, the key word method is used to produce sentence S_1.

132

iii. Cue method

The sentence S_1 is created by cue method, if each of sentences T_1 or T_2 contains one of the cue words (refer to Figure 5.10). According to the content of sentence T_1 and T_2, cue method was not used to produce sentence S_1.

iv. Location method

The sentence S_1 is created by location method if each of sentences T_1 or T_2 appears at the first or at the last of a paragraph. Therefore, according to the content of a text (refer to Appendix B.1), each of sentences T_1 and T_2 are not at the first or end of a paragraph. Thus, location method was not used to produce sentence S_1.

5.7 Runtime complexity analysis

Runtime-complexity of the RDSSIA can be affected by several factors such as the size (number of sentences) of the source text and summary text. The time complexity of RDSSIA is calculated as follows:

i. Let $T = (m + n)$ be the total number of sentence where m is the number of sentences in source text and n is the number of sentences in summary text. As the number of the sentences is equal to T, the time complexity of pre-processing stage is O (T).

ii. Intermediate-processing stage requires O $(n \times m)$ time to determine relevant sentences.

iii. According to the proposed algorithm, after the identifying relevant sentences, we need take one more step which is the post-processing. In this step all heuristic rules are applied to all relevant sentences to determine summarizing

133

strategies. The total complexity of this stage is O ($k \times p$) time where p and $k = 5$ indicate the number of extracted relevant sentences and the number of heuristic rules, respectively.

Thus, total time complexity becomes, O ($T \times k \times p \times n \times m$)..

5.8 Summary

In this chapter, we propose an algorithm to identify text relevancy and summarizing strategies, used to create each summary sentence. For this purpose, it should answer several questions (refer to section 5.1).

This algorithm responds to the aforementioned questions using the following major tasks:

- To measure sentence similarity using semantic sentences similarity and word order similarity.
- To identify original sentence that produced a *SS*.
- To identify various methods and summarizing strategies that were employed to produce each *SS*.

In the first task, the algorithm calculates the sentence similarity using composition of semantic and syntactic similarity measures between sentences. The semantic similarity measure relies on common words between two sentences; it assumes two sentences are more similar if they include more similar words. However, sometimes two sentences have different meaning, even they employ similar words. That is why we include word-order similarity to measure sentences similarity. On the other hand, there are also two sentences with similar meaning which do not share any common word; in this case we use semantic word similarity to find similar words between the two sentences. The second task is identifying original sentences have been used to create a *SS*. Finally, the

134

last task is used to determine the methods and summarizing strategies employed to create the *SS*.

CHAPTER 6

EXPERIMENTAL RESULTS AND EVALUATION

The aim of this chapter is to conduct an evaluation of RDSSIA. Firstly, we briefly describe the implementation of the system. An explanation of the two experiments carried out to evaluate RDSSIA is described next. In the first experiment, we evaluated the functionality of the algorithm to identify the summarizing strategies. In the second experiment, we measure the performance of the algorithm against human judgment to identify the summarizing strategies. Finally, we present the result of our evaluation experiment.

6.1 Implementation

To evaluate the performance of the proposed algorithm, we implemented it in C# with Windows 7 environment and tested it on the human summary, which is considered as the test data. Figure 6.1 shows the architecture of the RDSSI system. The main Stages of the system include:

- *Main User Interface* – allows the user to switch between the source text interface and the summary practice interface.

- *Source text interface* – allows the user to upload the source text for summary exercise.

- *Summary practice interface* – allows the user to write their summary for the source text.

- *RDSSIA* – applies the proposed algorithm to the student summary to determine summarizing strategies (refer to chapter 5).

- *Result interface* – displays the output of the system.

136

We also used some NLP tools and resources to support the implementation of RDSSIA: These tools and resources are:

- *Stop words* – includes a set of stop words (refer to Appendix D).

- *WordNet* – is a lexical database for English (Miller & Charles, 1991).

- *Cue words* – includes a set of cue words (refer to Table A.6 of Appendix A).

- *Part-Of-Speech Tagger*–assigns to each word its morphological category (Tsuruoka & Tsujii, 2005).
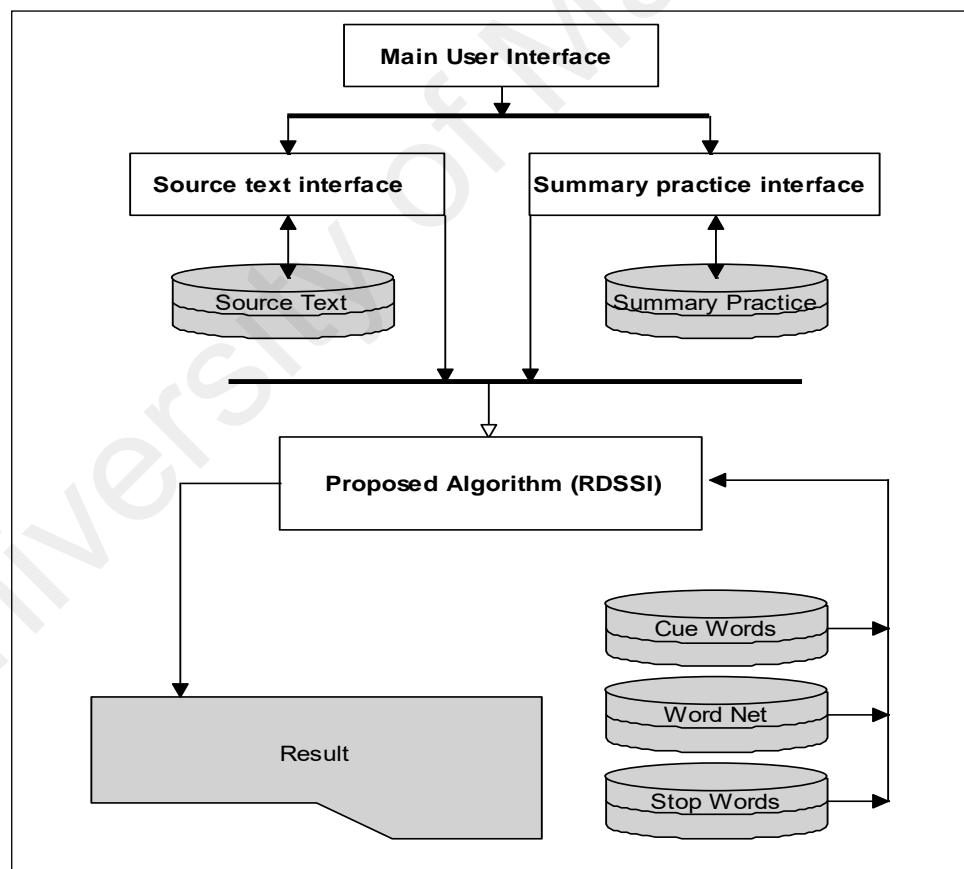
Figure 6.1: The Architecture of the RDSSI system

137

6.2 Experiment 1- Functionality of The System

In this experiment, we examine whether the system can detect the summarizing strategies employed by students in summary writing. We applied our system to students' summaries. However, if the system identifies the summarizing strategies employed by students to create each *SS*, we can conclude the system can determine summarizing strategies and can accomplish task. The procedure will explain in the next subsection.

6.2.1. Procedure

To evaluate our system for identifying summarizing strategies we used the students–generated summaries (refer to Appendix E) from a source text (refer to Appendix B.2).

The students' summaries and the source text were used to extract summarizing strategies by applying our proposed algorithm. Figure 6.1 shows how the result is obtained by the system. The input texts (source text and summary text) are first submitted to the main interface document, as shown in Figure 6.2. The input texts are then sent to the next Stage, *RDSSI Stage.* In this Stage, as we discussed in previous chapter several processes such as *pre–processing, intermediate–processing* and *post–processing* are applied toboth texts. Finally, the results, as shown in Figure 6.3 (A) and 6.3 (B) will be considered as the output of the system and will be presented to readers.
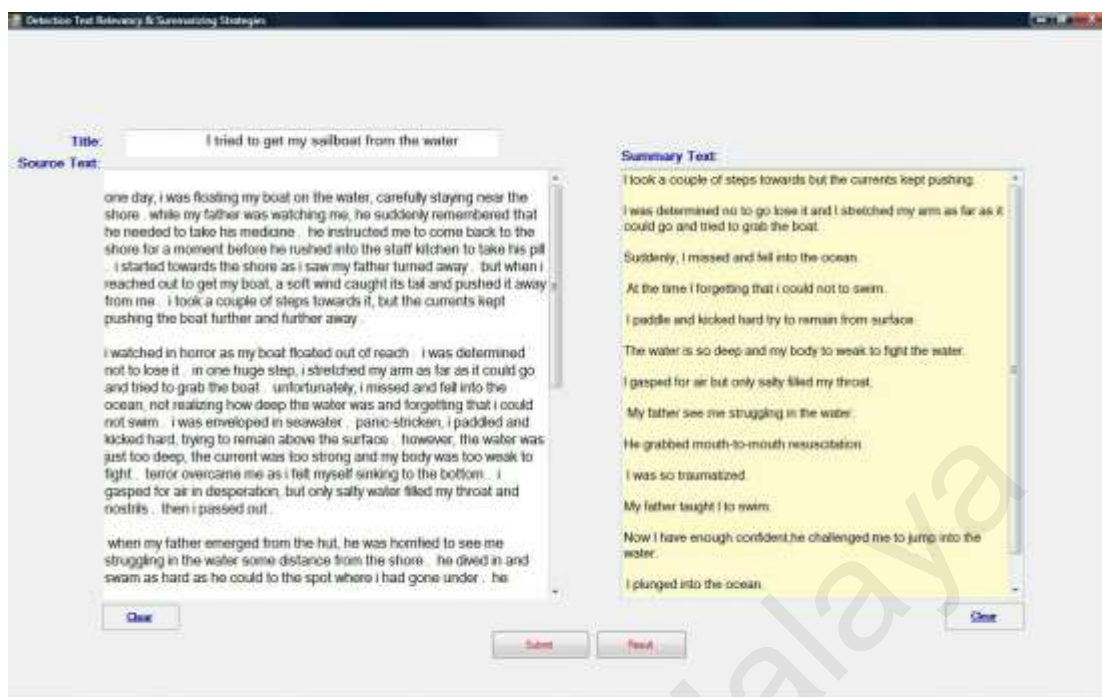
Figure 6.2: The main interface of the system

Figure 6.3 (A) shows the sentence that was used to create each *SS*, and also the methods and summarizing strategies that were employed by students to produce each sentence. Figure 6.3(B) displays the total number of methods and summarizing strategies employed to produce a *SS*. In addition, the system is able to assess students' summaries. Summary writing assessment can be classified into content and linguistic quality (Karen Sparck Jones & Galliers, 1996; Valenti, Neri, & Cucchiarelli, 2003). Content assessment identifies how much the information of the original text overlaps with the information in the summary text; on the other hand, linguistic quality assessment judges the accurate spelling and grammar of summaries, etc. Our system focuses on the content evaluation. It determines the content based similarity, as shown in Figure 6.3(B).
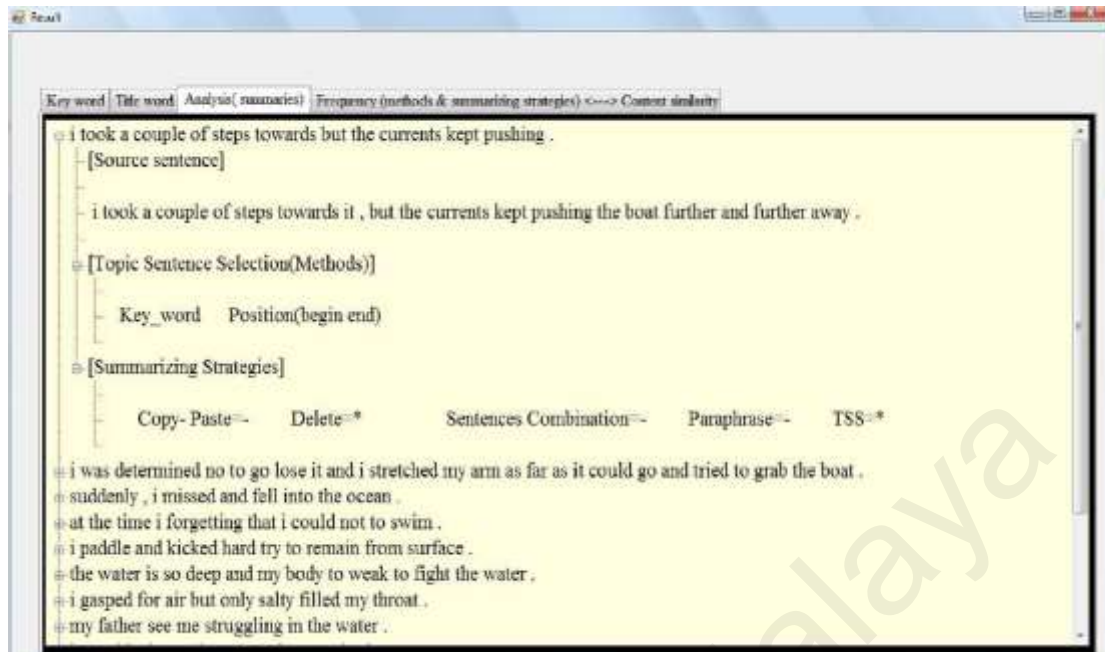
Figure 6.3 (A): The Result interface of the system
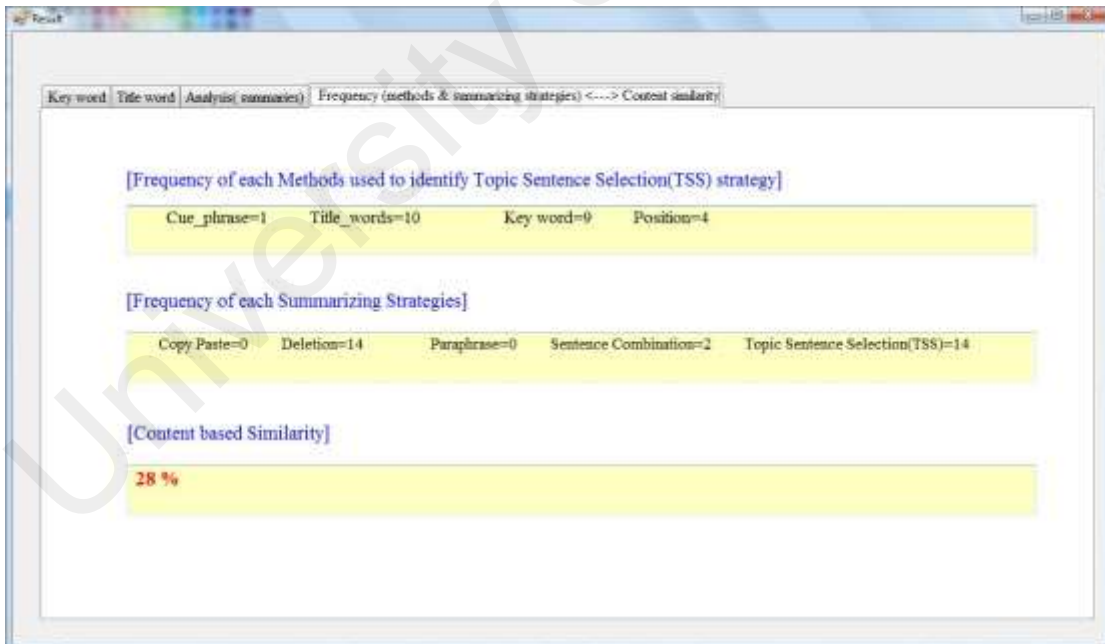
Figure 6.3 (B): The Result interface of the system

6.2.2. Result and Discussion

The summarizing strategies of students' summaries identified by RDSSIA are presented in Tables (E.1) - (E.3), (refer to Appendix E). As shown in these Tables, according to the heuristic rules (refer to chapter 4) the system identified the summarizing strategies such as deletion, sentence combination, paraphrase and topic sentence selection. The system also identified copy-verbatim, although as we discussed in Chapter 2. Besides the aforementioned strategies, there are four methods to identify TSS strategy: "*cue method*", "*title method*", "*keyword method*" and "*location method*", which can also be identified by the system.

Table 6.1 shows an example of the students' summaries identified by RDSSI system. For each student summary, Table 6.1 shows the number of each summary strategy that was used by a student. As shown in Table 6.1, 3 of the *SS* were produced using deletion strategy and the other 2 sentences were produced using the sentence combination strategy. The student also used T.S.S strategy to produce 3 sentences. The table also shows that the student used cue, title, key-word and location method to create the *SS*. In addition, the system can find the source sentences that were employed to create the *SS*. For example, the *SS:* "*A method to producing hypnosis is the hypnotist tries to obtain his subject co-operation by pointing out to him the advantages to be secured by the hypnosis*" is produced using two sentences as follows: i) "*There are many methods of producing hypnosis; indeed, almost every experienced hypnotist employs variations differing slightly from those of others*" and ii) "*The hypnotist tries to obtain his subject's co-operation by pointing out to him the advantages to be secured by the hypnosis, such as, for instance, the help in curing a nervous illness to be derived from the patient's remembering in the trance certain events which otherwise are inaccessible to his memory.*".

Based on the results in Table 6.1 and Figures 6.3 (A) and 6.3 (B) we have made the following main observation. The system can determine summarizing strategies in both levels, semantically and syntactically.

Table 6.1: Examples of summarizing strategies identified by RDSSI system

NO.	Source sentence(s)	Summary sentence	Summarizing strategies					Methods			
			D	S.C	P	TSS	CP	Cue	Tit	Loc	Key
1	"He will take up any suggestion the hypnotist puts forward and act on it to the best of his ability." (S$_{16}$)	"He will take up any suggestion."	√	-	-	√	-	-	-	-	√
2	"There are many methods of producing hypnosis; indeed, almost every experienced hypnotist employs variations differing slightly from those of others." (S$_1$) "The hypnotist tries to obtain his subject's co-operation by pointing out to him the advantages to be secured by the hypnosis, such as, for instance, the help in curing a nervous illness to be derived from the patient's remembering in the trance certain events which otherwise are inaccessible to his memory." (S$_3$)	"A method to producing hypnosis is the hypnotist tries to obtain his subject co-operation by pointing out to him the advantages to be secured by the hypnosis."	√	√	-	√	-	√	√	√	√
3	"It would not be true to say, however, that all suggestions are accepted, even in the very deepest trance." (S$_{21}$)	"It would not be true to say, however, that all suggestions are accepted, even in the very deepest trance."	-	-	-	-	√	-	-	-	-
4	"It is sometimes helpful to concentrate the subject's attention on some small bright object dangled just above eye-	"Then, concentrate the subject's attention by dangling small bright object above	√	√	-	√	-	√	-	-	√

142

| | | **Total:** | 3 | 2 | 0 | 3 | 1 | 1 | 2 | 1 | 2 |

Where,

D: Deletion, **SC**: Sentence combination, **TSS**: Topic Sentence Selection,

Inv: Invention, **CV**: Copy – verbatim, **G**: Generalization, **P**: Paraphrase,

Cue: cue method, **Tit**: Title method, **Loc**: Location method, **Key**: Key method

'√'= *Strategy or method is used to produce summary sentence.*

'-'= *Strategy or method is not used to produce summary sentence.*

6.3 Experiment 2 - Comparison with the human judgment

To examine the efficiency of the RDSSIA, we evaluated the performance of the RDSSIA against human judgment. To do this, we now explain our experiments on the students' summaries.

6.3.1. Precision, Recall and F–score

To evaluate the performance of the RDSSIA, evaluation metric includes precision, recall and F-measure, were used.

Precision, recall and F-score are the prevalent measures for evaluating a system (Fazlı, 2011; Manning, Raghavan, & Schütze, 2008). Precision is the fraction of selected items that are correct and recall is the fraction of correct items that are selected. In this study, the summarizing strategies identified by a human refer to a set of ideal items, and the strategies identified by an algorithm refer to a set of system items. Precision is used to

assess the fraction of the system items that the algorithm correctly identified and recall is used to assess the fraction of the ideal items that the algorithm identified. The precision is computed using equation (6.1). It is the division of identified summarizing strategies by RDSSIA and human expert over the number of summarizing strategies identified by the algorithm only. The recall is computed using equation (6.2). It is the division of identified summarizing strategies by RDSSIA and human expert intersection over the number of summarizing strategies identified by human expert.

$$\text{Pericision} = \frac{A}{A+B} \tag{6.1}$$

$$\text{Recall} = \frac{A}{A+C} \tag{6.2}$$

Where,

A = "The number of summarizing strategies identified by Algorithm and Human expert".

B = "The number of summarizing strategies identified by Algorithm only".

C= "The number of summarizing strategies identified by Human expert only".

There is an anti–correlation between precision and recall (Manning et al., 2008). Hence, we used F-score, a statistical measure that merges both precision and recall, Eq. 6.3.

$$\text{F} - \text{measure} = \frac{1}{\alpha \times \frac{1}{P} + (1-\alpha)\frac{1}{R}} = \frac{(\beta^2+1)P \times R}{\beta^2 \times P + R} \tag{6.3}$$

Where $\beta^2 = \frac{1-\alpha}{\alpha}$, $\alpha \in [0\,,1]$, and $\beta^2 \in [0\,,\infty]$. If $(\beta > 1)$, it means the precision has more priority. If $(\beta < 1)$, it means the recall has more priority. If $\beta = 1$ the precision and recall are assumed to have equally priority Eq. (6.4).

$$\text{F} - \text{measure} = \frac{2 \times P \times R}{P+R} \tag{6.4}$$

144

Where *P* is precision and *R* is recall.

6.3.1.1. Procedure

To evaluate our algorithm for identifying summarizing strategies, we used two data sets, training data and testing data, using students' summaries. All students' summaries (refer to Appendix C) and an article (refer to Appendix B.1) are considered as a training data. All students' summaries (refer to Appendix E) and an article (refer to Appendix B.2) are considered as a testing data. We need a gold standard data in our experiment. For this purpose, we asked two experts to identify the summarizing strategies for each SS. These summarizing strategies are determined using the procedure as mentioned in section 3.3.2

The summarizing strategies determined by human expert as an example, are presented in Table 6.2, which is extracted from Appendix C. In particular, for each student summary, the first column shows the corresponding source sentences, the second column presents the SS_s; and finally the last column displays the summarizing strategies. It is worth mentioning, we also used the single-document summarization datasets provided by DUC (http://duc.nist.gov) to evaluate our proposed algorithm, for more information, refer to work published by (Abdi et al., 2016).

Table 6.2: Summarizing strategies identified by Human expert

Original sentence	Summary sentence	Summarizing strategy
"Panic-stricken, I paddled and kicked hard, trying to remain above the surface." (S_{12})	"I paddled and kicked hard trying to remain surface."	▪ Deletion ▪ Title word ▪ T.S.S
"However, the water was just too deep, the current was too strong and my body was too weak to fight." (S_{13})	"The water was just too deep, strong and my body was too weak to fight."	▪ Deletion ▪ Cue ▪ Title word ▪ T.S.S
"Then I passed out. "(S_{16})	"Then I passed out."	▪ Location ▪ T.S.S ▪ Copy-verbatim
"When my father emerged from the hut, he was horrified to see me struggling in the water some distance from the shore." (S_{17})	"My father was horrified to see me struggling in the water from the shore."	▪ Deletion ▪ Key word ▪ Location ▪ Title word ▪ T.S.S
"He dived in and swam as hard as he could to the spot where I had gone under." (S_{18})	"He dived and swam as hard as he could to spot where I had gone."	▪ Deletion ▪ Key word ▪ T.S.S
"He repeatedly dived under the water, frantically searching for my body." (S_{19})	"He dived under water searching for my body. "	▪ Deletion ▪ Cue ▪ Title word ▪ T.S.S
"He grabbed it and yanked my head out of the water." (S_{21})	"He grabbed it and yanked my head out. "	▪ Deletion ▪ Location ▪ Title word ▪ T.S.S
"In the days that followed, I was so traumatized that I would not go near the water." (S_{25})	"I was traumatized."	▪ Deletion ▪ Key word ▪ Location ▪ Title word ▪ T.S.S
"In the following weeks, my father taught me how to swim." (S_{29})	"My father taught me to swim."	▪ Deletion ▪ Key word ▪ T.S.S
"In the following weeks, my father taught me how to swim. (S_{35}) As my father proudly looked on, I knew I had overcome my fear." (S_{29})	"My father proudly looked on I know how to swim."	▪ Deletion ▪ Sentence combination ▪ Key word ▪ Location ▪ T.S.S

6.3.1.2. Parameter setting

The RDSSIA requires a weighting parameter (λ) (refer to equation (5.5)) for weighting the significance between semantic information and syntactic information. The parameter was determined using training data. We applied the RDSSIA on the current data set. Table 6.3 displays summarizing strategies identified by RDSSIA (third column) as an example.

Table 6.3: Summarizing strategies identified by RDSSIA and Human expert (training data)

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
"My father dived and swarms as hard as he could to the spot where i had gone under."	- Deletion - Key word - T.S.S	- Deletion - Sentence combination - Key word - T.S.S
"I was determined not to lose it."	- Location - T.S.S - Copy-verbatim	Copy-verbatim
"I felt myself sinking to the bottom and my father save me."	- Deletion - Sentence combination - Key word - T.S.S - Invention	- Deletion - Sentence combination - Paraphrase - Key word - Location - T.S.S

We evaluate RDSSIA for each (λ) between 0.1 to 0.9 with a step of 0.1. Table 6.4 presents our experimental results achieved by using various λ values. We evaluated the results in terms of precision, recall and f-measure using equations (6.1), (6.2) and (6.4) respectively. It is worth noting. The Ave_p, Ave_R, and $Ave_{F-score}$ denote the average

precision, the average recall and the average f-measure of total training data for each λ value.

$$Ave_p = \frac{\sum_{i=1}^{N} \text{Pericision}_i}{N} \tag{6.5}$$

$$Ave_R = \frac{\sum_{i=1}^{N} \text{Recall}_i}{N} \tag{6.6}$$

$$Ave_{\text{F-measure}} = \frac{\sum_{i=1}^{N} \text{F-measure}_i}{N} \tag{6.7}$$

Where, N is the number of students' summaries.

According to the results, as shown in Table 6.4, we found that the best performance is achieved when $\lambda=0.8$. This λ produced the scores for three metrics as follows: 0.8060 (precision), 0.6918 (recall), 0.7445 (F-measure).

Table 6.4: Comparison between human and RDSSIA against various λ values

Weighting (λ)	Ave_P	Ave_R	$Ave_{\text{F-measure}}$
0.1	0.6249	0.5385	0.5785
0.2	0.6394	0.5343	0.5821
0.3	0.6404	0.5770	0.6070
0.4	0.6501	0.5990	0.6235
0.5	0.6882	0.5853	0.6326
0.6	0.7316	0.6955	0.7131
0.7	0.7441	0.7194	0.7315
0.8	0.8060	0.6918	0.7445
0.9	0.7659	0.6454	0.7005

To confirm the results, as shown in Table 6.4, we measured the performance of the RDSSIA against human judgment using unused data set, testing data. Table 6.5 shows

summarizing strategies identified by RDSSIA (third column) as an example, which is extracted from Appendix E.

Table 6.5: Summarizing strategies identified by RDSSIA and Human expert (testing data)

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	RDSSIA
"There are methods of producing hypnosis; example subject will ask to lie down on the couch."	- Deletion - Key word - Title word - Location - Cue - T.S.S - Sentence combination	- Deletion - Key word - Title word - Location - Cue - T.S.S - Sentence combination
"The hypnotized can be very useful and helpful for some people, for example to release their tension and some hypnotized has been misused."	Invention	- Deletion - Key word - Cue - T.S.S - Sentence combination
"Light trance will induce."	- Deletion - Key word - Cue - T.S.S	- Deletion - Key word - Cue - T.S.S
"Having induced a reasonably deep hypnotic trance in our subject, what types of phenomena can be elicited."	- Copy-verbatim - Key word - Location - T.S.S	- Copy-verbatim - Key word - Location - T.S.S
"Such practices are not to be encouraged because they go counter to the ideal of human dignity."	- Deletion - Key word - Location - Cue - T.S.S	- Deletion - Key word - Location - Cue - T.S.S

6.3.1.3. Results and Analysis

We determine the number of summarizing strategies identified by the algorithm and human expert (A), the number of summarizing strategies identified by algorithm only

(B), and the number of summarizing strategies identified by human only (C). Then, the equations of precision, recall and F-score are applied to obtain the values for each summary. Table 6.6 presents an example of the results (refer to Appendix E; Table E.4).

Table 6.6: Precision, Recall and F-score

Summary	A	B	C	Precision	Recall	F-score
1	4	2	2	0.67	0.67	0.67
2	5	0	2	1.00	0.71	0.83
3	6	0	2	1.00	0.75	0.86
4	5	0	2	1.00	0.71	0.83
5	7	1	1	0.88	0.88	0.88

Table E.4 shows the algorithm achieved an average of 87% precision, 83% recall and 85% F-measure.

It did not attain a high percentage for the precision, recall and F-score in comparison to human judgment due to various reasons, such as:

   i.   The RDSSIA is not able to identify some of the summarizing strategies such as generalization and invention.

   ii.   The algorithm and human identified the TSS strategy using the cue method. However, there is no standard list of cue words.

   iii.   In RDSSIA, WordNet has been employed to compute SSBW. However, the WordNet is not able to cover all words to compute SSBW. Obviously, this shortcoming has a negative effect on the performance of the RDSSIA.

   iv.   The RDSSIA cannot distinguish between an active and a passive sentence. To tackle this problem the algorithm needs a NLP technique such as Semantic Role

Labeling (SRL) or shallow semantic parsing. Shallow semantic parsing is used to determine semantic role (i.e. verb, subject, object, place, time and etc.) for each word inside the sentences. For example, consider the sentence (*A*: "*Father likes his child.*") and two original sentences (*B*: "*Child likes his father.*"; *C*: "*Child is liked by his father.*"), the verb of this sentence is *likes*, *Father* is the subject and *child* is the object. In sentence *B* also, the verb of this sentence is *likes*, *Child* is the subject and *Father* is the object. To calculate the degree of similarity between two sentences *A* and *B*, we can compare subject with subject, verb with verb, etc. Therefore it becomes clear in comparison between two sentences (*A* and *B*) based on the semantic allocation for each term inside the sentence, since the subject and object of two sentences are different, the meaning of two sentence is also different

6.3.2. Accuracy

We also experimented on dataset with the method, called accuracy. The accuracy of an algorithm is the portion of its outputs that are correct. For our experiment, the accuracy was computed according to the equation (6.8).

$$\text{Accuracy} = \frac{T\_Sen}{N} \times 100 \qquad (6.8)$$

Where,

T_Sen = total number of summary sentences which have the same strategies identified by RDSSIA and human expert.

N= whole summaries sentences.

151

6.3.2.1. Procedure

In this experiment, we used the testing dataset, as mentioned in section 6.3.1.1. First, the summarizing strategies for each summary text are determined by RDSSIA and human expert. Then, the intersection and total number of summarizing strategies detected by both RDSSIA and human expert are determined. The result of our experiment is shown in Table E.5 (refer to Appendix E). Table 6.7 displays an example of the results.

Table 6.7: No. of the same summarizing strategies identified by RDSSIA and Human expert

Summary	Number of summary sentences which have same summarizing strategies	Number of summary sentences in each summary
1	2	4
2	3	5
3	5	7
4	3	5
5	7	7

6.3.2.2. Results

From the Table 6.7, it is found that the accuracy of our algorithm compared to human judgement is 82%.

$$\text{Accuracy} = \frac{T\_Sen}{N} \times 100$$

$$= \frac{297}{363} \times 100$$

$$= 82\%$$

In other words, the algorithm and human expert found 82% similar summarizing strategies. Based on the three main reasons (ii, iii and iv) as discussed in section 6.3.1.3, the algorithm did not achieve the high accuracy measure (accuracy rate > 82%).

6.4 Summary

This chapter presents the evaluation of the algorithm and shows its potential when tested on students' summaries for identifying summarizing strategies. It started by describing the implementation of the algorithm. We then explained the functionality of the system. Finally, we evaluated our algorithm. In our experiment, the algorithm achieved an average 87% precision, 83% recall, 85% F-score and 82% accuracy.

The precision, recall, F–score and accuracy rate evaluation show that the proposed algorithm achieved acceptable results. Although the results are good in testing students' summaries, we detected several problems in ensuring the algorithm to perform well. To make it more applicable, we focused on improving the performance of the algorithm. Currently, the algorithm can identify the deletion, paraphrase, topic sentence selection and sentence combination and topic sentence selection strategy, but in the future the result will be improved when the algorithm is able to identify others strategies such as generalization and invention, and also, the external resources, cue–word list and stop–word list are revised.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

Since summarization is an important tool for improving reading comprehension (Chiu et al., 2013; Westby et al., 2010), it has attracted interest from teachers to teach summary writing through direct instruction (Westby et al., 2010). In direct instruction, teachers need to possess some information, such as what summarizing strategies used by students, the ability of students to use summarizing strategies, and the students' weakness in summarizing. To collect this information manually is difficult as it is a highly time consuming task. To tackle these problems, Computer-Assisted Assessment (CAA), which has garnered much interest in recent years, is one of the methods that can be used to assist teachers.

Most of the previous systems focus on content coverage. Hence, we aim is to develop an efficient algorithm into an automated summarization assessment system that can be used to identify the strategies. In order to achieve this goal, many literatures from various domains have been reviewed, such as automatic text summarization, approaches to text summarization, summarization assessment, summary assessment techniques and general rules for producing a summary.

The contributions, conclusion and future work of the work are presented in the next sections.

7.1 Summary of the Contributions

This thesis makes four contributions to the field of Natural Language Processing and related field of automated summarization assessment. We summarize these contributions below:

- **C1: Sentence similarity computation model**. It addresses the text relevance detection problem. Text relevance detection is a necessary prerequisite of the summarizing strategies identification. This model applies both semantic relations between words and their syntactic composition for computing sentences similarity measure.

- **C2: Identifying summarizing strategies**. We formulate a set of heuristic rules into an algorithm to identify summarizing strategies at semantic and syntactic levels.

- **C3: Development an algorithm based on contributions Sentence similarity computation model and Identifying summarizing strategies at semantic and syntactic levels**.
  We contribute an algorithm in automated summarization assessment that takes linguistic measure to identify text relevancy and summarizing strategies. This algorithm has proved to be extremely robust and successful. It is also easy to deploy.

- **C4: A helpful tool for teachers and a learning environment for students**.
  The proposed algorithm is a helpful tool for teachers and students. It assists the teachers finding out the students' ability in use of summarizing strategies. Moreover, it helps students to improve their skills in summary writing.

7.2 Conclusion

In this thesis, we have described an algorithm, called RDSSIA, to identify students' summarizing strategies. This algorithm is based on linguistic measure. We show that the RDSSIA deals with both source text and summary text which in the first step finds all relevant sentences for each summary sentence from source text and then identifies the summarizing strategies which have been employed to create each summary sentence.

It is worth noting that the proposed algorithm does not need deeper linguistic processing than just tokenization, part-of-speech tagging and the lexical database. This helps in keeping the portability across languages that shallow NLP techniques allow.

To evaluate the RDSSIA we conducted two experiments using students' summary sentences. The first experiment is to examine the functionality of the algorithm to identify the summarizing strategies by applying the algorithm onto students' summary sentences. The obtained result of the experiment shows that the algorithm is able to identify summarizing strategies and methods such as deletion, sentence combination, paraphrase, topic sentence selection, copy-verbatim, cue method, title method, keyword method and location method. Due to the result, we conclude that the algorithm is able to identify summarizing strategies and method syntactically and semantically. In second experiment we evaluated the performance of the algorithm with the human judgments. The results are as follow: an average of 87% precision, 83% recall, 85% F-score and 82% accuracy for students' summaries.

7.3 Future works

Given the merits the research has, some limitations have also been identified. These limitations will be addressed in our future work. Some general limitations are summarized as follows:

- We aim to work on algorithms for identifying other summarizing strategies such as generalization and invention.

- The algorithm employed WordNet to calculate semantic similarity measure between two words. However, the WordNet is not able to cover all words. Hence, to tackle this problem, in addition to WordNet, we aim to used other knowledge resources.

- We are confident that identifying the passive and active sentences before comparing two sentences can improve performance of the proposed algorithm.

- Preparing a comprehensive cue-word list for identifying cue method and topic sentence selection strategy in order to improve the performance of the proposed algorithm.

- Further enhancement of RDSSIA by reducing the algorithm runtime with a parallel programming and adding additional functionality to the algorithm.

We also consider the following works as future works.

- To compare our experimental results with results obtained with other algorithms. In other words, we compare our algorithm with the other proposed algorithms to identify summarizing strategies. The existing system used different data set and evaluation measure in their experiment. Moreover, the developed system is not available. Therefore, we re-examined the existing algorithms in future.

- To apply the SRDS and SSCS of the proposed algorithm to other applications. The Sentences Relevance Detection Stage (SRDS) and Sentence Similarity Computation Stage (SSCS) can be applied to other application such as Non Factoid Question Answering (NFQA) to identify relevant sentence with the user-

query from a document; and also summarization based on the user-query to produce a summary base on the user-query.

- To assess the summary without a reference summary. The common way to assess the content of the summaries is to compare them with a reference summary, which is a hard and expensive task. Much effort is required to have a corpus of texts and their corresponding summaries. To resolve this problem, the proposed algorithm through a few modifications can be used for assessing the content of the summary, which a reference summary is no longer necessary, and the algorithm takes the original text and summary text as its input to assess the summary.

Reference

Achananuparp, P., Hu, X., & Shen, X. (2008). The evaluation of sentence similarity measures *Data Warehousing and Knowledge Discovery* (pp. 305-316): Springer.

Alemany, L. A., de Lingüıstica General, D., Masalles, I. C., & Cirera, L. P. (2005). Representing discourse for automatic text summarization via shallow NLP techniques. *Unpublished PhD thesis, Universitat de Barcelona*.

Alguliev, R. M., Aliguliyev, R. M., & Isazade, N. R. (2013). CDDS: Constraint-driven document summarization models. *Expert Systems with Applications, 40*(2), 458-465.

Alguliev, R. M., Aliguliyev, R. M., & Mehdiyev, C. A. (2011). Sentence selection for generic document summarization using an adaptive differential evolution algorithm. *Swarm and Evolutionary Computation, 1*(4), 213-222.

Aliguliyev, R. M. (2009). A new sentence similarity measure and sentence based extractive technique for automatic text summarization. *Expert Systems with Applications, 36*(4), 7764-7772.

Allen, J. (1987). Natural language understanding. *Phiosophy Documentation Center*. doi:http://philpapers.org/rec/ALLNLU

Allen, M. (2004). *Assessing academic programs*: Boston: Anker Publishing.

Alonso, L. (2005). *Representing discourse for automatic text summarization via shallow NLP techniques*. Tesis doctoral. Barcelona: Universitat de Barcelona.

Alonso, L., Castellón, I., Climent, S., Fuentes, M., Padró, L., & Rodrıguez, H. (2004). Approaches to text summarization: Questions and answers. *Inteligencia Artificial, 8*(2), 22.

AlSanie, W., Touir, A., & Mathkour, H. (2005). *Towards an infrastructure for Arabic text summarization using rhetorical structure theory*. M. Sc. Thesis, King Saud University, Riyadh, Saudi Arabia.

Alyousef, H. S. (2006). Teaching reading comprehension to ESL/EFL learners. *Journal of language and learning, 5*(1), 63-73.

Barry, A. L. (2002). Reading strategies teachers say they use. *Journal of adolescent and adult literacy, 46*(2), 132-143.

Bartlett, F., & Burt, C. (1933). Remembering: a study in experimental and social psychology. *British Journal of Educational Psychology, 3*(2), 187-192.

Baumann, J. F. (1984). The effectiveness of a direct instruction paradigm for teaching main idea comprehension. *Reading Research Quarterly*, 93-115.

Bawakid, A., & Oussalah, M. (2008). *A semantic summarization system: University of Birmingham at TAC 2008*. Paper presented at the Proceedings of the First Text Analysis Conference.

Baxendale, P. B. (1958). Machine-made index for technical literature: an experiment. *IBM Journal of research and development, 2*(4), 354-361.

Bean, T. W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Literacy Research, 16*(4), 297-306.

Brandow, R., Mitze, K., & Rau, L. F. (1995). Automatic condensation of electronic publications by sentence selection. *Information Processing & Management, 31*(5), 675-685.

Brown, A. L., Campione, J. C., & Day, J. D. (1981). Learning to learn: On training students to learn from texts. *Educational researcher*, 14-21.

Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of verbal learning and verbal behavior, 22*(1), 1-14.

Cai, X., & Li, W. (2011). A spectral analysis approach to document summarization: clustering and ranking sentences simultaneously. *Information Sciences, 181*(18), 3816-3827.

Cai, X., Li, W., & Zhang, R. (2014). Enhancing diversity and coverage of document summaries through subspace clustering and clustering-based optimization. *Information Sciences, 279*, 764-775.

Casazza, M. E. (1993). Using a model of direct instruction to teach summary writing in a college reading class. *Journal of Reading*, 202-208.

Chang, K.-E., Sung, Y.-T., & Chen, I.-D. (2002). The effect of concept mapping to enhance text comprehension and summarization. *The Journal of Experimental Education, 71*(1), 5-23.

Chatterjee, N., & Sahoo, P. K. (2015). Random Indexing and Modified Random Indexing based approach for extractive text summarization. *Computer Speech & Language, 29*(1), 32-44.

Chen, C. C., & Chen, M. C. (2012). TSCAN: A content anatomy approach to temporal topic summarization. *Knowledge and Data Engineering, IEEE Transactions on, 24*(1), 170-183.

Chiu, C.-H., Wu, C.-Y., & Cheng, H.-W. (2013). Integrating reviewing strategies into shared electronic note-taking: Questioning, summarizing and note reading. *Computers & Education, 67*(4), 229-238.

Cho, Y. (2012). *Teaching Summary Writing through Direct Instruction to Improve Text Comprehension for Students in ESL/EFL Classroom.* University of Wisconsin-River Falls.

Choi, F. Y., Wiemer-Hastings, P., & Moore, J. (2001, 2001). *Latent semantic analysis for text segmentation.* Paper presented at the In Proceedings of EMNLP.

Conroy, J. M., Schlesinger, J. D., & O'Leary, D. P. (2011). Nouveau-rouge: A novelty metric for update summarization. *Computational Linguistics, 37*(1), 1-8.

Cunningham, J. W. (1982). Generating interactions between schemata and text. *New inquiries in reading research and instruction, 42*, 47.

De Marneffe, M.-C., MacCartney, B., & Manning, C. D. (2006). *Generating typed dependency parses from phrase structure parses.* Paper presented at the Proceedings of LREC.

Duke, N. K., & Pearson, P. D. (2008). Effective practices for developing reading comprehension. *Journal of education*, 107-122.

Edmundson, H. P. (1969). New methods in automatic extracting. *Journal of the ACM (JACM), 16*(2), 264-285.

Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Intell. Res.(JAIR), 22*(1), 457-479.

Erwin, T. D. (1991). Assessing Student Learning and Development: A Guide to the Principles, Goals, and Methods of Determining College Outcomes.

Fan, Y.-C. (2010a). The Effect of Comprehension Strategy Instruction on EFL Learners' Reading Comprehension. *Asian Social Science, 6*(8).

Fan, Y.-C. (2010b). The Effect of Comprehension Strategy Instruction on EFL Learners' Reading Comprehension. *Asian Social Science, 6*(8), P19.

Farzindar, A., Rozon, F., & Lapalme, G. (2005). *CATS a topic-oriented multi-document summarization system at DUC 2005.* Paper presented at the Proc. of the 2005 Document Understanding Workshop (DUC2005).

Fattah, M. A., & Ren, F. (2009). GA, MR, FFNN, PNN and GMM based models for automatic text summarization. *Computer Speech & Language, 23*(1), 126-144.

Fazlı, C. (2011). *TEXT SUMMARIZATION USING LATENT SEMANTIC ANALYSIS.* (Doctoral dissertation, MIDDLE EAST TECHNICAL UNIVERSITY).

Ferreira, R., de Souza Cabral, L., Lins, R. D., e Silva, G. P., Freitas, F., Cavalcanti, G. D., . . . Favaro, L. (2013). Assessing sentence scoring techniques for extractive text summarization. *Expert Systems with Applications, 40*(14), 5755-5764.

Fialding, L., & Pearson, P. D. (1994). Synthesis of research reading comprehension: What works. *Educational Leadership, 51*, 62-62.

Franzke, M., & Streeter, L. A. (2006). Building student summarization, writing and reading comprehension skills with guided practice and automated feedback. *Highlights From Research at the University of Colorado, A white paper from Pearson Knowledge Technologies*(2).

Fraser, B. (1999). What are discourse markers? *Journal of pragmatics, 31*(7), 931-952.

Friend, R. (2001). Effects of strategy instruction on summary writing of college students. *Contemporary Educational Psychology, 26*(1), 3-24.

Galgani, F., Compton, P., & Hoffmann, A. (2014). HAUSS: Incrementally building a summarizer combining multiple techniques. *International Journal of Human-Computer Studies, 72*(7), 584-605.

Garner, R. (1982). Efficient text summarization: Costs and benefits. *The Journal of Educational Research*, 275-279.

Gholamrezazadeh, S., Salehi, M. A., & Gholamzadeh, B. (2009). A comprehensive survey on text summarization systems. *Proceedings of CSA, 9*, 1-6.

Glavaš, G., & Šnajder, J. (2014). Event graphs for information retrieval and multi-document summarization. *Expert Systems with Applications, 41*(15), 6904-6916.

Goldstein, J., Mittal, V., Carbonell, J., & Kantrowitz, M. (2000). *Multi-document summarization by sentence extraction.* Paper presented at the Proceedings of the 2000 NAACL-ANLPWorkshop on Automatic summarization-Volume 4.

Graham, S., & Hebert, M. (2010). *Writing to read: Evidence for how writing can improve reading: A report from Carnegie Corporation of New York*: Carnegie Corporation of New York.

Guido, B., & Colwell, C. G. (1987). A rationale for direct instruction to teach summary writing following expository text reading. *Literacy Research and Instruction, 26*(2), 89-98.

Gupta, V., & Lehal, G. S. (2010a). A Survey of Text Summarization Extractive Techniques. *Journal of Emerging Technologies in Web Intelligence, 2*(3).

Gupta, V., & Lehal, G. S. (2010b). A survey of text summarization extractive techniques. *Journal of Emerging Technologies in Web Intelligence, 2*(3), 258-268.

Hahn, U., & Romacker, M. (2001). *The SYNDIKATE text Knowledge base generator.* Paper presented at the Proceedings of the first international conference on Human language technology research.

Hare, V. C., & Borchardt, K. M. (1984). Direct instruction of summarization skills. *Reading Research Quarterly*, 62-78.

Hartley, J., & Trueman, M. (1983). The effects of headings in text on recall, search and retrieval. *British Journal of Educational Psychology, 53*(2), 205-214.

He, Li, F., Shao, W., Chen, J., & Ma, L. (2008). *A New Feature-Fusion Sentence Selecting Strategy for Query-Focused Multi-document Summarization.* Paper presented at the Advanced Language Processing and Web Information Technology, 2008, ALPIT'08. .

He, Y., Hui, S. C., & Quan, T. T. (2009). Automatic summary assessment for intelligent tutoring systems. *Computers & Education, 53*(3), 890-899.

Hedge, T. (2001). *Teaching and learning in the language classroom* (Vol. 106): Oxford University Press Oxford,, UK.

Hill, M. (1991). Writing Summaries Promotes Thinking and Learning across the Curriculum: But Why Are They so Difficult to Write? *Journal of Reading*, 536-539.

Hirschberg, J., & Litman, D. (1993). Empirical studies on the disambiguation of cue phrases. *Computational Linguistics, 19*(3), 501-530.

Hovy, E., & Lin, C.-Y. (1998). *Automated text summarization and the SUMMARIST system.* Paper presented at the Proceedings of a workshop on held at Baltimore, Maryland: October 13-15, 1998.

Hovy, E., & Marcu, D. (2005). Automated text summarization. *The Oxford handbook of computational linguistics*, 583-598.

Huang, S.-c. (2006). Reading English for academic purposes–What situational factors may motivate learners to read? *System, 34*(3), 371-383.

Huba, M. E., & Freed, J. E. (2000). Learner centered assessment on college campuses: Shifting the focus from teaching to learning. *Community College Journal of Research and Practice, 24*(9), 759-766.

Idris, N., Baba, S., & Abdullah, R. (2009). A Summary Sentence Decomposition Algorithm for Summarizing Strategies Identification. *Computer and Information Science, 2*(1), P200.

Inderjeet, M. (2001). *Summarization Evaluation: An Overview.* Paper presented at the Proceedings of the NTCIR Workshop.

Jing, H. (2001). *Cut-and-paste text summarization.* Columbia University.

Jing, H., Barzilay, R., McKeown, K., & Elhadad, M. (1998). *Summarization evaluation methods: Experiments and analysis.* Paper presented at the AAAI Symposium on Intelligent Summarization.

Jing, H., & McKeown, K. R. (2000). *Cut and paste based text summarization.* Paper presented at the Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference.

John, A., & Wilscy, M. (2015). Vertex Cover Algorithm Based Multi-document Summarization Using Information Content of Sentences. *Procedia Computer Science, 46*, 285-291.

Johnson, N. S. (1983). What do you do if you can't tell the whole story? The development of summarization skills. *Children's language, 4*(4), 315-383.

Jones, K. S. (1999). Automatic summarizing: factors and directions. *Advances in automatic text summarization*, 1-12.

Jones, K. S., & Galliers, J. R. (1996). *Evaluating natural language processing systems: An analysis and review* (Vol. 1083): Springer.

Kamhi-Stein, L. (1993). Summarization, Notetaking, and Mapping Techniques: Lessons for L2 Reading Instruction.

Kanejiya, D., Kumar, A., & Prasad, S. (2003). *Automatic evaluation of students' answers using syntactically enhanced LSA.* Paper presented at the Proceedings of the HLT-NAACL 03 workshop on Building educational applications using natural language processing-Volume 2.

Karbalaei, A., & Rajyashree, K. (2009). The Impact of Summarization Strategy Training on University ESL Learners" Reading Comprehension. *The International Journal of Language Society and Culture, 30*(3), 41-53.

Karbalaei, A., & Rajyashree, K. (2010). The Impact of Summarization Strategy Training on University ESL Learners" Reading Comprehension. *International Journal of Language Society and Culture, Issue, 30*(3), 41-53.

Kashef, S. H., Damavand, A., & Viyani, A. (2012). Strategies-Based ESP Instruction (SBI) of Reading Comprehension: Male vs. Female Students. *International Journal of Education, 4*(2), p171-p180.

Kazantseva, A., & Szpakowicz, S. (2010). Summarizing short stories. *Computational Linguistics, 36*(1), 71-109.

Khoshsima, H., & Nia, M. R. (2014). Summarizing strategies and writing ability of Iranian intermediate EFL students.

Kintsch, Patel, V. L., & Ericsson, K. A. (1999). The role of long-term working memory in text comprehension. *Psychologia, 42*(4), 186-198.

Kintsch, Steinhart, D., Stahl, G., LSA Research Group, L. R. G., Matthews, C., & Lamb, R. (2000). Developing summarization skills through the use of LSA-based feedback. *Interactive Learning Environments, 8*(2), 87-109.

Kintsch, & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review, 85*(5), 363.

Knott, A. (1996). A data-driven methodology for motivating a set of coherence relations.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*: Cambridge University Press.

Kruengkrai, C., & Jaruskulchai, C. (2003). *Generic text summarization using local and global properties of sentences.* Paper presented at the Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on.

Kupiec, J., Pedersen, J., & Chen, F. (1995). *A trainable document summarizer.* Paper presented at the Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval.

Landauer, T. K. (2002). On the computational basis of learning and cognition: Arguments from LSA. *Psychology of learning and motivation, 41*, 43-84.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review, 104*(2), 211.

Landauer, T. K., Laham, D., Rehder, B., & Schreiner, M. E. (1997). *How well can passage meaning be derived without using word order? A comparison of Latent Semantic Analysis and humans.* Paper presented at the Proceedings of the 19th annual meeting of the Cognitive Science Society.

Lee, Y.-J. (2010). *Copying and summarizing: Possible tools to develop English reading and writing for university students of different proficiency levels in Korea.* Indiana University.

Lemaire, B., Mandin, S., Dessus, P., & Denhière, G. (2005). Computational cognitive models of summarization assessment skills. (2).

León, J. A., & Carretero, M. (1995). Intervention in comprehension and memory strategies: Knowledge and use of text structure. *Learning and instruction, 5*(3), 203-220.

León, J. A., Olmos, R., Escudero, I., Cañas, J. J., & Salmerón, L. (2006). Assessing short summaries with human judgments procedure and latent semantic analysis in narrative and expository texts. *Behavior research methods, 38*(4), 616-627.

Li, Y., McLean, D., Bandar, Z. A., O'shea, J. D., & Crockett, K. (2006). Sentence similarity based on semantic nets and corpus statistics. *Knowledge and Data Engineering, IEEE Transactions on, 18*(8), 1138-1150.

Lin. (2004). *Rouge: A package for automatic evaluation of summaries.* Paper presented at the Text Summarization Branches Out: Proceedings of the ACL-04 Workshop.

Lin, & Hovy, E. (2003). *Automatic evaluation of summaries using n-gram co-occurrence statistics.* Paper presented at the Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1.

Lin, D. (1998). *An information-theoretic definition of similarity.* Paper presented at the ICML.

Lloret, E. (2012). Text summarisation based on human language technologies and its applications. *Procesamiento del lenguaje natural, 48*, 119-122.

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development, 2*(2), 159-165.

Mani, I., Klein, G., House, D., Hirschman, L., Firmin, T., & Sundheim, B. (2002). SUMMAC: a text summarization evaluation. *Natural Language Engineering, 8*(01), 43-68.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (Vol. 1): Cambridge university press Cambridge.

Marzano, R. J. (2003). *What works in schools: Translating research into action*: ASCD.

Marzano, R. J. (2006). *Classroom assessment & grading that work*: ASCD.

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*: ASCD.

McNeil, J., & Donant, L. (1982). Summarization strategy for improving reading comprehension. *New inquiries in reading research and instruction, 215*, 219.

Mendoza, M., Bonilla, S., Noguera, C., Cobos, C., & León, E. (2014). Extractive single-document summarization based on genetic operators and guided local search. *Expert Systems with Applications, 41*(9), 4158-4169.

Mihalcea, R., & Tarau, P. (2004). *TextRank: Bringing order into texts*.

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and cognitive processes, 6*(1), 1-28.

Mishra, R., Bian, J., Fiszman, M., Weir, C. R., Jonnalagadda, S., Mostafa, J., & Del Fiol, G. (2014). Text summarization in the biomedical domain: a systematic review of recent research. *Journal of biomedical informatics, 52*, 457-467.

Mohler, M., Bunescu, R., & Mihalcea, R. (2011). *Learning to grade short answer questions using semantic similarity measures and dependency graph alignments.* Paper presented at the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1.

Pakzadian, M., & Rasekh, A. E. (2013). The Effects of Using Summarization Strategies on Iranian EFL Learners' Reading Comprehension. *English Linguistics Research, 1*(1), p118.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and instruction, 1*(2), 117-175.

Palomba, C. A., & Banta, T. W. (1999). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education. Higher and Adult Education Series*: ERIC.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). *BLEU: a method for automatic evaluation of machine translation.* Paper presented at the Proceedings of the 40th annual meeting on association for computational linguistics.

Pearson, P. D., & Fielding, L. (1991). Comprehension instruction. *Handbook of reading research, 2*, 815-860.

Pembe, F. C., & Güngör, T. (2007). *Automated Querybiased and Structure-preserving Text Summarization on Web Documents.* Paper presented at the Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications, İstanbul.

Pérez, Alfonseca, E., & Rodrıguez, P. (2004). *Upper bounds of the BLEU algorithm applied to assessing student essays.* Paper presented at the Proceedings of the 30th international association for educational assessment (IAEA) conference.

Pérez, Bandera, E. E., León, I. R., & Cerván, R. L. (2003). Application of a Written Composition and Reading Comprehension Program in Primary Education Students.

Pérez, Gliozzo, A. M., Strapparava, C., Alfonseca, E., Rodríguez, P., & Magnini, B. (2005). *Automatic Assessment of Students' Free-Text Answers Underpinned by the Combination of a BLEU-Inspired Algorithm and Latent Semantic Analysis.* Paper presented at the FLAIRS Conference.

Pressley, M. (1998). Comprehension strategies instruction. *Literacy for all: Issues in teaching and learning*, 113-133.

Qazvinian, V., Hassanabadi, L. S., & Halavati, R. (2008). Summarising text with a genetic algorithm-based sentence extraction. *International Journal of Knowledge Management Studies, 2*(4), 426-444.

Qazvinian, V., & Radev, D. R. (2008). *Scientific paper summarization using citation summary networks.* Paper presented at the Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1.

Rankel, P. A., Conroy, J. M., & Schlesinger, J. D. (2012). Better metrics to automatically predict the quality of a text summary. *Algorithms, 5*(4), 398-420.

Riedl, C. (2011). How to get the most from summarizing strategies? *TEXAS SCHOOL IMPROVEMENT CONFERENCE 2011.Texas: unpublished*, p5.

Rinehart, S. D., Stahl, S. A., & Erickson, L. G. (1986). Some effects of summarization training on reading and studying. *Reading Research Quarterly*, 422-438.

Saggion, H., & Poibeau, T. (2013). Automatic text summarization: Past, present and future *Multi-source, Multilingual Information Extraction and Summarization* (pp. 3-21): Springer.

Sarker, A., Mollá, D., & Paris, C. (2013). An approach for query-focused text summarisation for evidence based medicine *Artificial Intelligence in Medicine* (pp. 295-304): Springer.

Selinger, B. (1993). Starting at the Top: Using Hierarchical Structure To Train College Students To Summarize.

Shareghi, E., & Hassanabadi, L. S. (2008). *Text summarization with harmony search algorithm-based sentence extraction.* Paper presented at the Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology.

Sobh, I., Darwish, N., & Fayek, M. (2006). *A trainable Arabic Bayesian extractive generic text summarizer.* Paper presented at the Proceedings of the Sixth Conference on Language Engineering ESLEC.

Steinberger, J., & Ježek, K. (2012). Evaluation measures for text summarization. *Computing and Informatics, 28*(2), 251–275.

Steinberger, J., Poesio, M., Kabadjov, M. A., & Ježek, K. (2007). Two uses of anaphora resolution in summarization. *Information Processing & Management, 43*(6), 1663-1680.

Swales, J. M., Christine B. Feat. (1994). How to Write a Summary. *Academic Writing for Graduate Students, Essential Tasks and Skills. Ann Arbor: U Michigan P. unpublished*105-130.

Swan, M. (2005). *Practical english usage* (Vol. 95): Oxford University Press Oxford.

Taylor, K. K. (1983). Can college students summarize? *Journal of Reading*, 524-528.

Taylor, K. K. (1986). Summary writing by young children. *Reading Research Quarterly*, 193-208.

Teufel, S., & Moens, M. (1997). *Sentence extraction as a classification task.* Paper presented at the Proceedings of the ACL.

Tian, Y., Li, H., Cai, Q., & Zhao, S. (2010). *Measuring the similarity of short texts by word similarity and tree kernels.* Paper presented at the Information Computing and Telecommunications (YC-ICT), 2010 IEEE Youth Conference on.

Tsuruoka, Y., & Tsujii, J. i. (2005). *Bidirectional inference with the easiest-first strategy for tagging sequence data.* Paper presented at the Proceedings of the conference on human language technology and empirical methods in natural language processing.

Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An overview of current research on automated essay grading. *Journal of Information Technology Education: Research, 2*(1), 319-330.

Wang, D., & Li, T. (2012). Weighted consensus multi-document summarization. *Information Processing & Management, 48*(3), 513-523.

Westby, C., Culatta, B., Lawrence, B., & Hall-Kenyon, K. (2010). Summarizing expository texts. *Topics in Language Disorders, 30*(4), 275-287.

Wiemer-Hastings, P., & Wiemer, P. (2000). *Adding syntactic information to LSA.* Paper presented at the Proceedings of the 22nd Annual Meeting of the Cognitive Science Society.

Wiemer-Hastings, P., & Zipitria, I. (2001). *Rules for syntax, vectors for semantics.* Paper presented at the Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society.

Willis, J. (1981). *Teaching English through English* (Vol. 8): Harlow: Longman.

Winograd, P. N. (1984). Strategic difficulties in summarizing texts. *Reading Research Quarterly*, 404-425.

Wormeli, R. (2005). *Summarization in any subject: 50 techniques to improve student learning*: Association for Supervision and Curriculum Development.

Xie, S., & Liu, Y. (2008). *Using corpus and knowledge-based similarity measure in maximum marginal relevance for meeting summarization.* Paper presented at the Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on.

Xie, S., & Liu, Y. (2010). Improving supervised learning for meeting summarization using sampling and regression. *Computer Speech & Language, 24*(3), 495-514.

Xie, S., Liu, Y., & Lin, H. (2008). *Evaluating the effectiveness of features and sampling in extractive meeting summarization.* Paper presented at the Spoken Language Technology Workshop, 2008. SLT 2008. IEEE.

Yang, G., Chen, N.-S., Sutinen, E., Anderson, T., & Wen, D. (2013). The effectiveness of automatic text summarization in mobile learning contexts. *Computers & Education, 68*(4), 233-243.

Yang, G., Wen, D., & Sutinen, E. (2013). *A Contextual Query Expansion Based Multi-document Summarizer for Smart Learning.* Paper presented at the Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference on.

Zafarani, P., & Kabgani, S. (2014). Summarization Strategy Training and Reading Comprehension of Iranian ESP Learners. *Procedia-Social and Behavioral Sciences, 98*, 1959-1965.

Zhang, J., Sun, L., & Zhou, Q. (2005). *A cue-based hub-authority approach for multi-document text summarization.* Paper presented at the Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE'05. Proceedings of 2005 IEEE International Conference on.

Zhao, X., & Tang, J. (2010). *Query-focused Summarization Based on Genetic Algorithm.* Paper presented at the 2010 International Conference on Measuring Technology and Mechatronics Automation.

Zipitria, I., Arruarte, A., & Elorriaga, J. A. (2010). *Automatically Grading the Use of Language in Learner Summaries.* Paper presented at the Proceedings of the 18th International Conference on Computers in Education, Putrajaya, Malaysia.

Zipitria, I., Elorriaga, J. A., Arruarte, A., & de Ilarraza, A. D. (2004). *From human to automatic summary evaluation.* Paper presented at the Intelligent Tutoring Systems.

PhD Related Publications

Journals:

1. Abdi, A., Idris, N., Alguliyev, R. M., & Aliguliyev, R. M. (2016). An Automated Summarization Assessment Algorithm for Identifying Summarizing Strategies. Plos One, 11(1).

2. Abdi, A., Idris, N., Alguliev, R. M., & Aliguliyev, R. M. (2015). Automatic summarization assessment through a combination of semantic and syntactic information for intelligent educational systems. Information Processing & Management, 51(4), 340-358.

3. Abdi, A., Idris, N., Alguliyev, R. M., & Aliguliyev, R. M. (2015). PDLK: Plagiarism detection using linguistic knowledge. Expert Systems with Applications.

4. Abdi, A., Idris, N., Alguliyev, R. M., & Aliguliyev, R. M. (2015). Query-based multi-documents summarization using linguistic knowledge and content word expansion. Soft Computing, 1-17.

5. Abdi, S. A., & Idris, N. (2014). An Analysis on Student-Written Summaries: Automatic Assessment of Summary Writing. International Journal of Enhanced Research in Science Technology & Engineering, 3(1), 466-472.

Proceedings:

1. Abdi, A., & Idris, N. (2014). Automated summarization assessment system: quality assessment without a reference summary. Paper presented at the The International Conference On Advances In Applied Science And Environmental Engineering - ASEE 2014.

2. Abdi, A., Idris, N. (2015). Relevance Detection And Summarizing Strategies Identification Using Linguistic Measures. At: Universiti Teknologi MARA (UiTM)- kuala lumpur, Malaysia, Conference: NATIONAL INVENTION, INNOVATION, DESIGN & RESEARCH 2015, Affiliation: University of Malaya, DOI: 10.13140/RG.2.1.1214.0883.

Appendix A

Lexicon of Discourse Markers

Table A.1: Lexicon of Discourse Markers

Additionally	Earlier	in the future
a striking difference	equally important	in the meantime
above all	especially	in the second place
actually	final finally	in the third place
after a short time	finally	in this case
after all	following	in truth
after that following	for all that	indeed
again	for one thing indeed	it may well be
against	Formerly	just then
all at once	frequently	last
all of the time	from this point earlier	last of all
all the while	Further	later on
also exactly	Further to this	least of all
although	hence	like
although this may be true	I accept	little by little
always	I allow	matching
and	I believe	more and more
another	I grant you	more importantly
as a consequence of	I suppose,	moreover
as a matter of fact	I think	most of all
as a matter of fact	I'm afraid	namely
as a result	identically	naturally
as a rule	immediately	never
as an example for	in accord with	nevertheless
as illustrated by	in addition	next
as long as	in addition to	no doubt
as revealed by	in brief	on the next occasion
as soon as	in comparison	not at all
as stated	in conclusion	not long ago
at first	In consequence	now
at last	In contrast	obviously
at present	in fact	of course
at the beginning	in general	of little difference
at the end	in most cases	of major interest
at the same time	in much	

Table A.2: Lexicon of Discourse Markers

nonetheless	to explain	alternatively
probability	to explain	Although
quickly	to finish	altogether
Recently	to highlight	always assuming that
Regardless of the fact that	to oppose	and
repeatedly	to outline	and yet
resembling	to paraphrase	anyway
same as	to put it in another way	apart from
second secondly	to recapitulate	as
seldom	to rephrase	as well
similar to	to retell	as a consequence
similarly	to review	as a corollary
so far	to some extent	As a matter of fact
some of the time	to stress	As a result
sometimes	to sum up	as first
soon	to summarize	as long as
soon later	to tell the truth	As revealed
still	today	as soon as
suppose that specifically	tomorrow	As well as
that is to say	too	as with
the climax of	unfortunately	Assuming that
the following week	until	at any rate
the main issue is	usually	at first blush
the major point	when	at first sight
the next time then	while	at first view
therefore	while it is true	at the moment when
third	without a doubt	at the outset
thus	yet	at the same time
to add to that	about	because
to complete	after all	By
to conclude	after that	by comparison
to confess	afterward	by contrast
to culminate	again	by means of
to demonstrate	All in all	by the same token
broadly speaking	in one way	On the other hand
by now	in short	opposite
chiefly	in simpler terms	parallel to
currently	in the case of	periodically

to differ from	for the simple reason	by the way
to emphasize	for this reason	certainly
to end	Frankly	Consequently
again and again	further	in the beginning
all the same	furthermore	in the case of
also	generally	in the end
considering	gradually	In the event of
characteristic	hence	in the first place
clearly	given that	in the meantime
conversely	henceforth	In the same way
despite	Honestly	in this way
Despite the fact that	however	in turn
despite this fact	most of all	Incidentally
despite that	if	Indeed
either…or	if ever	infrequently
distinction	if not	Initially
due to	if only	Insofar as
during	if so	Instance
earlier	illustrated by	Instantly
equally	important of all	Instead of
equally important	in a different vein	it is true that
else	in contrast	it might seem that
essentially	in actual fact	just as
even so	in effect	largely
even then	in addition	Lastly
even though	in another way	Later
eventually	in fact	later on
every time	in any case	least of all
except	in as much as	least of all
following	in more often	Moreover
for	in my opinion	most of all
for a start	in order to	Mostly
For example	In spite of the fact that	much later
For instance	in sum	much sooner
for the moment	in that case	neither…nor
but	in opposition to versus	Once
by all means	in other words	once more
by contrast	in relation to	once upon time

Table A.4: Lexicon of Discourse Markers

neither…nor	rather	Thereby
nevertheless	reciprocally	Therefore
next	regardless of that	Therefore
no doubt	second	Third
no doubt	secondly	Thirdly
nonetheless	significantly	this time
not	similarly	Though
not because	simply because	Thus
not only	simultaneously	to be sure
not that	simultaneously	To be truthful
notwithstanding that	since	to begin with
notably	slowly	to continue
now	so	to illustrate
now that	So long as	to put it differently
obviously	so that	to reconsider
occasionally	specially	to repeat
of course	specifically	to show
often	starting with initially	to start with
on the other hand	still	to sum up
on condition that	Subsequently	to summarise
on one hand	such as	to take an example
on one side	such that	to the degree that
on the assumption that	Suddenly	to the extent that
on the contrary	summarising	Too
on the ground that	summing up	Towards
on the one hand	suppose	Ultimately
on the one side	suppose that	Undoubtedly
on the other hand	supposing that	Unless
on the other side	sure enough	Unlike
on the whole	surely absolutely	Unquestionably
once	temporarily	Until
precisely	the most necessary	Wherein
presumably because	the most significant	While
previously	the same way	with respect to

Table A.5: Lexicon of Discourse Markers

although this	yet	despite (doing) this/that
		in comparison (with/to this/that)
Rarely	contrary to this/that	
without question	conversely	in contrast (with/to this/that)
yesterday	by the same token	in spite of (doing) this/that
instead (of (doing) this/that)	I mean	on top of it all
rather (than (do) this/that)	more to the point	to cap it all off
better yet	what is more	
directly	in the end	Presently
during	in the first place	Previously
primarily	then	without exception
Provided that	then again	Thereafter
put another way	in short	Naturally
for one thing	in spite of	neither is it the case
otherwise	that's why	what's more
over and over	the chief	When
overall	the fact is	Whenever
plainly	the fact is that	Whereas
once again	thanks to	usually
once more	that is	we might say
or else	that is to say	well now
originally	that subsequently	what is more
except insofar as	in follows that	let us assume
extremely	in case	Like
far from	in comparison	Likewise
Finally	in conclusion	may be true
first	in other words	Meanwhile
first of all	in particular	Merely
firstly	in might appear that	merely because
to admit	Above all	because of
to agree	Accordingly	Because of this
to be exact	Actually	Before
to be sure	Additionally	Besides
to begin with	admittedly	Besides
to bring to an end	after all	Between
to clarify	after	But
to close	After a while	but also

Table A.6: Lexicon of Discourse Markers

Accordingly	in that case	in general
all things considered	in this case	in particular
as a conclusion	insomuch	in short
as a consequence	it can be concluded that	in summary
as a consequence of	last of all	in effect
as a logical	of course	to summarize
as a result	on that condition	given that
because	our investigation	in case
because of that	result	that's why
Because of this	significantly	However
conclusion	summarize	Important
consequence	the paper describe	in additional
consequently	thereby	in conclusion
end	therefore	In consequence
eventually	thereupon	to recapitulate
finally	though	to sum up
For this reason	Thus	To conclude
hardly	to an end	To end
hence	to bring	to complete
hereupon		to close

Appendix B

SAMPLE OF TEXT

B.1 Sample of a summary writing exercise

One day, I was floating my boat on the water, carefully staying near the shore. While my father was watching me, he suddenly remembered that he needed to take his medicine. He instructed me to come back to the shore for a moment before he rushed into the staff kitchen to take his pill. I started towards the shore as I saw my father turned away. But when I reached out to get my boat, a soft wind caught its tail and pushed it away from me. I took a couple of steps towards it, but the currents kept pushing the boat further and further away.

I watched in horror as my boat floated out of reach. I was determined not to lose it. In one huge step, I stretched my arm as far as it could go and tried to grab the boat. Unfortunately, I missed and fell into the ocean, not realizing how deep the water was and forgetting that I could not swim. I was enveloped in seawater. Panic-stricken, I paddled and kicked hard, trying to remain above the surface. However, the water was just too deep, the current was too strong and my body was too weak to fight. Terror overcame me as I felt myself sinking to the bottom. I gasped for air in desperation, but only salty water filled my throat and nostrils. Then I passed out.

When my father emerged from the hut, he was horrified to see me struggling in the water some distance from the shore. He dived in and swam as hard as he could to the spot where I had gone under. He repeatedly dived under the water, frantically searching for my body.

Then his hand brushed against my hair. He grabbed it and yanked my head out of the water. He pulled me to shore and started mouth-to-mouth resuscitation. When I recovered, I opened my eyes to see my father crying. I had never before, and have never since, seen him cry so hard.

In the days that followed, I was so traumatized that I would not go near the water. My father was worried that the incident would scare me for life. One day, he took me to the edge of the water and urged me to go in with him. Though scared, I went in because I trusted him. In the following weeks, my father taught me how to swim. When I was confident enough, he took me to a jetty and challenged me to jump into the water. I

knew that the water was deep, but I was not looking at it. I was looking at my father, who showed great faith and confidence in me. I plunged into the ocean and swam back to shore. I was overjoyed. As my father proudly looked on, I knew I had overcome my fear.

B.2 Original Text on hypnosis

There are many methods of producing hypnosis; indeed, almost every experienced hypnotist employs variations differing slightly from those of others. Perhaps the most common method is something along these lines. The hypnotist tries to obtain his subject's co-operation by pointing out to him the advantages to be secured by the hypnosis, such as, for instance, the help in curing a nervous illness to be derived from the patient's remembering in the trance certain events which otherwise are inaccessible to his memory. The patient is reassured about any possible dangers he might suspect to be present in hypnosis, and he may also be told (quite truthfully) that it is not a sign of instability or weakness to be capable of being put in a hypnotic trance, but that, quite on the contrary, a certain amount of intelligence and concentration on the part of the subject is absolutely essential.

Next, the subject is asked to lie down on a couch, or sit in an easy-chair. External stimulation is reduced to a minimum by drawing the curtains and excluding, as far as possible, all disruptive noises. It is sometimes helpful to concentrate the subject's attention on some small bright object dangled just above eye-level, thus forcing him to look slightly upwards. This leads quickly to a fatigue of the eye-muscles, and thus facilitates his acceptance of the suggestion that he is feeling tired and that his eyes are closing. The hypnotist now begins to talk to the subject in a soft tone of voice, repeating endlessly suggestions to the effect that the subject is feeling drowsy, getting tired, that his eyes are closing, that he is falling into a deep sleep, that he cannot hear anything except the hypnotist's voice, and so on and so forth. In a susceptible subject, a light trance is thus induced after a few minutes, and the hypnotist now begins to deepen this trance and to test the reactions of the subject by giving suggestions which are more and more difficult of execution. Thus, he will ask the subject to clasp his hands together, and tell him that it is impossible for him to separate his hands again. The subject, try as he may, finds, to his astonishment, that he cannot in actual fact pull his hands apart. Successful suggestions of this kind are instrumental in deepening the hypnotic trance until, finally, in particularly good subjects; all the phenomena which will be discussed presently can be elicited.

Having induced a reasonably deep hypnotic trance in our subject, what types of phenomena can be elicited? The first and most obvious one, which, indeed, may be responsible in large measure for all the others, is a tremendous increase in the subject's suggestibility. He will take up any suggestion the hypnotist puts forward and act on it to

the best of his ability. Suggest to him that he is a dog, and he will go down on all fours and rush around the room barking and yelping. Suggest to him that he is Hitler, and he will throw his arms about and produce an impassioned harangue in an imitation of the raucous tones of the Fuhrer! This tremendous increase in suggestibility is often exploited on the stage to induce people to do foolish and ridiculous acts. Such practices are not to be encouraged because they go counter to the ideal of human dignity and are not the kind of way in which hypnosis ought to be used; nevertheless, they must be mentioned because it is probably phenomena such as these which are most familiar to people from vaudeville acts, from reading the papers, and so forth.

It would not be true to say, however, that all suggestions are accepted, even in the very deepest trance. This is particularly true when a suggestion is made which is contrary to the ethical and moral conceptions held by the subject. A well-known story may be quoted to illustrate this. Charcot, the great French neurologist, whose classes at one time were attended by Freud, was lecturing on hypnosis and was demonstrating the phenomena of the hypnotic trance on a young girl of eighteen. When she had been hypnotized deeply he was called away, and handed over the demonstration to one of his assistants. This young man, lacking the seriousness of purpose so desirable in students of medicine, even French ones, suggested to the young lady that she should remove her clothes. She immediately awakened from her trance, slapped his face, and flounced out of the room, very much to his discomfiture.

Appendix C

ANALYSIS

C.1 Analysis on Summary Sentences

- **Some samples**

Sample #1

Original sentence	Summary sentence	Summarizing strategy
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S_{12})	After that, I paddled and kicked hard, trying to remain above the surface.	▪ Deletion ▪ Title word ▪ T.S.S
I gasped for air in desperation, but only salty water filled my throat and nostrils. (S_{15})	I gasped for air in desperation; the salty water filled my throat and nostrils.	▪ Title word ▪ T.S.S ▪ Copy-verbatim
He repeatedly dived under the water, frantically searching for my body. (S_{19})	He frantically is searching for my body.	▪ Deletion ▪ Cue ▪ Location ▪ T.S.S
He pulled me to shore and started mouth-to-mouth resuscitation. (S_{22})	He pulled me to shore and started mouth-to-mouth resuscitation.	▪ Key word ▪ T.S.S ▪ Cope-verbatim
One day, he took me to the edge of the water and urged me to go in with him. (S_{27}) In the following weeks, my father taught me how to swim. (S_{29})	He took me to the edge of the water, and he taught me how to swim.	▪ Deletion ▪ Sentence combination ▪ Key word ▪ Title word ▪ T.S.S
When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S_{30})	He took me to a jetty and challenged me to jump into the water.	▪ Deletion ▪ Key word ▪ T.S.S

Sample #2

Original sentence	Summary sentence	Summarizing strategy
I started towards the shore as I saw my father turned away. (S_4)	I started toward the shore.	▪ Deletion ▪ Key word ▪ T.S.S

Original sentence	Summary sentence	Summarizing strategy
But when I reached out to get my boat, a soft wind caught its tail and pushed it away from me. (S5)	I reached to get my boat; a soft wind caught its tail and pushed it away.	▪ Deletion ▪ Key word ▪ T.S.S
I took a couple of steps towards it, but the currents kept pushing the boat further and further away. (S6)	The currents kept pushing the boat further and further away.	▪ Deletion ▪ Key word ▪ Cue ▪ Location ▪ T.S.S
I was determined not to lose it. (S8)	I was determined not to lose it.	▪ Location ▪ T.S.S ▪ Cope-verbatim
In one huge step, I stretched my arm as far as it could go and tried to grab the boat. (S9)	In one huge step, I stretched my arm as far as it could go and tried to grab the boat.	▪ Key word ▪ T.S.S ▪ Cope-verbatim
I was enveloped in seawater. (S11)	I was enveloped in seawater.	▪ Cope-verbatim
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S12)	Panic-stricken, I paddled and kicked hard, trying to remain above the surface.	▪ Key word ▪ T.S.S ▪ Cope-verbatim
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S12) However, the water was just too deep, the current was too strong and my body was too weak to fight. (S13)	However stricken I paddled and kicked hard, trying to remain above the surface.	▪ Deletion ▪ Sentence combination ▪ Cue ▪ Title word ▪ T.S.S
However, the water was just too deep, the current was too strong and my body was too weak to fight. (S13)	The water was just too deep, and current strong my body was too weak to fight.	▪ Deletion ▪ Title word ▪ Cue ▪ T.S.S
I was looking at my father, who showed great faith and confidence in me. (S32)	I showed great faith and confidence.	▪ Deletion ▪ Key word ▪ T.S.S

Sample # 3

Original sentence	Summary sentence	Summarizing strategy
I was determined not to lose it. 8 In one huge step, I stretched my arm as far as it could go and tried to	I was determined not to go lose it and I stretched my arm as far as it could go and tried	▪ Deletion ▪ Sentence combination ▪ Location

grab the boat. (S₉)	to grab the boat.	▪ Title word ▪ Key word ▪ T.S.S ▪ Copy-verbatim
Unfortunately, I missed and fell into the ocean, not realizing how deep the water was and forgetting that I could not swim. (S₁₀)	Suddenly, I missed and fell into the ocean.	▪ Deletion ▪ Title word ▪ Key word ▪ T.S.S
Unfortunately, I missed and fell into the ocean, not realizing how deep the water was and forgetting that I could not swim. (S₁₀)	At the time I forgetting that i could not to swim.	▪ Deletion ▪ Title word ▪ Key word ▪ T.S.S
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S₁₂)	I paddle and kicked hard try to remain from surface.	▪ Deletion ▪ Title word ▪ T.S.S
However, the water was just too deep, the current was too strong and my body was too weak to fight. (S₁₃)	The water is so deep and my body too weak to fight the water.	▪ Deletion ▪ Paraphrase ▪ Title word ▪ Cue ▪ T.S.S
I gasped for air in desperation, but only salty water filled my throat and nostrils. (S₁₅)	I gasped for air but only salty filled my throat.	▪ Deletion ▪ Title word ▪ T.S.S
When my father emerged from the hut, he was horrified to see me struggling in the water some distance from the shore. (S₁₇)	My father sees me struggling in the water.	▪ Deletion ▪ Title word ▪ Key word ▪ Location ▪ T.S.S
He grabbed it and yanked my head out of the water. (S₂₁) He pulled me to shore and started mouth-to-mouth resuscitation. (S₂₂)	He grabbed mouth-to-mouth resuscitation.	▪ Deletion ▪ Sentence combination ▪ Title word ▪ Key word ▪ location ▪ T.S.S
In the days that followed, I was so traumatized that I would not go near the water. (S₂₅)	I was so traumatized.	▪ Deletion ▪ Title word ▪ Key word ▪ location ▪ T.S.S
In the following weeks, my father taught me how to swim. (S₂₉)	My father taught I to swim.	▪ Deletion ▪ Key word ▪ T.S.S
When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S₃₀)	Now I have enough confident, he challenged me to jump into the water.	▪ Deletion ▪ Title word ▪ Key word ▪ T.S.S

Original sentence	Summary sentence	Summarizing strategy
I plunged into the ocean and swam back to shore. (S$_{33}$)	I plunged into the ocean.	▪ Deletion ▪ Location ▪ Key word ▪ T.S.S
As my father proudly looked on, I knew I had overcome my fear. (S$_{35}$)	I had overcome my fear.	▪ Deletion ▪ Location ▪ Key word ▪ T.S.S

Sample # 4

Original sentence	Summary sentence	Summarizing strategy
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S$_{12}$)	I paddled and kicked hard trying to remain above the surface.	▪ Deletion ▪ Title word ▪ T.S.S
However, the water was just too deep, the current was too strong and my body was too weak to fight. (S$_{13}$)	However the water was just to deep.	▪ Deletion ▪ Title word ▪ Cue ▪ T.S.S
Terror overcame me as I felt myself sinking to the bottom. (S$_{14}$) I gasped for air in desperation, but only salty water filled my throat and nostrils. (S$_{14}$)	I felt myself sinking to the bottom only salty water filled my throat and nostrils.	▪ Deletion ▪ Sentence combination ▪ Title word ▪ T.S.S
When my father emerged from the hut, he was horrified to see me struggling in the water some distance from the shore. (S$_{17}$)	Me struggling in the water some distance from the shore.	▪ Deletion ▪ Title word ▪ Key word ▪ location ▪ T.S.S
He dived in and swam as hard as he could to the spot where I had gone under. (S$_{18}$)	He dived in and swam as hard as he could to the spot had gone under.	▪ Key word ▪ T.S.S ▪ Copy-verbatim
When I recovered, I opened my eyes to see my father crying. (S$_{23}$)	I opened my eyes to see my father crying.	▪ Deletion ▪ Key word ▪ T.S.S
I had never before, and have never since, seen him cry so hard. (S$_{24}$)	Have never since, seen him cry so hard.	▪ Deletion ▪ Key word ▪ Location ▪ T.S.S
In the days that followed, I was so traumatized that I would not go near the water. (S$_{25}$)	I was traumatised that I would not go near the water.	▪ Deletion ▪ Key word ▪ Title word ▪ Location ▪ T.S.S
In the following weeks, my father taught me how to swim. (S$_{29}$)	My father taught me how to swim.	▪ Deletion ▪ Key word ▪ T.S.S

184

Original sentence	Summary sentence	Summarizing strategy
When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S₃₀)	I was confident enough he took me to a jetty and challenged me jump into water.	▪ Key word ▪ T.S.S ▪ Copy-verbatim

Sample # 5

Original sentence	Summary sentence	Summarizing strategy
Unfortunately, I missed and fell into the ocean, not realizing how deep the water was and forgetting that I could not swim. (S₁₀)	Then, I fell into the ocean.	▪ Deletion ▪ Title word ▪ T.S.S
Unfortunately, I missed and fell into the ocean, not realizing how deep the water was and forgetting that I could not swim. (S₁₀)	The water is deep and I could not swim.	▪ Deletion ▪ Title word ▪ T.S.S
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S₁₂)	I paddled and kick hard to remain above surface.	▪ Deletion ▪ Title word ▪ T.S.S
However, the water was just too deep, the current was too strong and my body was too weak to fight. (S₁₃)	The current was strong and my body was weak.	▪ Deletion ▪ Title word ▪ Cue ▪ T.S.S
I gasped for air in desperation, but only salty water filled my throat and nostrils. (S₁₅)	I gasped for air but only salty my throat and nostrils.	▪ Deletion ▪ Title word ▪ T.S.S
When my father emerged from the hut, he was horrified to see me struggling in the water some distance from the shore. (S₁₇)	Father was horrified to see me struggling.	▪ Deletion ▪ Title word ▪ Location ▪ Key word ▪ T.S.S
He repeatedly dived under the water, frantically searching for my body. (S₁₉)	He repeatedly dived under the water searching me.	▪ Deletion ▪ Title word ▪ Location ▪ Key word ▪ T.S.S
He pulled me to shore and started mouth-to-mouth resuscitation. (S₂₂)	He pulled me and started mouth-to-mouth resuscitation.	▪ Deletion ▪ Key word ▪ T.S.S
When I recovered, I opened my eyes to see my father crying. (S₂₃)	I opened my eyes to see my father crying.	▪ Deletion ▪ Key word ▪ T.S.S
I had never before, and have never since, seen him cry so hard. (S₂₄)	I had never before see him cry.	▪ Deletion ▪ Location ▪ T.S.S

Original sentence	Summary sentence	Summarizing strategy
In the days that followed, I was so traumatized that I would not go near the water. (S$_{25}$)	One day I would not go near water.	▪ Deletion ▪ Title word ▪ Location ▪ Key word ▪ T.S.S
In the following weeks, my father taught me how to swim. (S$_{29}$)	My father taught me how to swim.	▪ Deletion ▪ Key word ▪ T.S.S
When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S$_{30}$)	He challenged me to jump into the water.	▪ Deletion ▪ Title word ▪ Key word ▪ T.S.S
I plunged into the ocean and swam back to shore. (S$_{33}$) As my father proudly looked on, I knew I had overcome my fear. (S$_{35}$)	I plunged into the ocean and swam back looked and overcome fear.	▪ Deletion ▪ Sentence combination ▪ Location ▪ Key word ▪ T.S.S

Sample # 6

Original sentence	Summary sentence	Summarizing strategy
However, the water was just too deep, the current was too strong and my body was too weak to fight. (S$_{13}$)	My body was too weak to fight the current.	▪ Deletion ▪ Title word ▪ Cue ▪ T.S.S
Terror overcame me as I felt myself sinking to the bottom. (S$_{14}$)	And then I felt myself at the bottom.	▪ Deletion
When my father emerged from the hut, he was horrified to see me struggling in the water some distance from the shore. (S$_{17}$) When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S$_{30}$)	After that, my father emerged jump into the water when he see me struggling in the water.	▪ Deletion ▪ Sentence combination ▪ Title word ▪ Location ▪ Key word ▪ T.S.S
He repeatedly dived under the water, frantically searching for my body. (S$_{19}$)	He is searching me.	▪ Deletion ▪ Title word ▪ Location ▪ T.S.S
He pulled me to shore and started mouth-to-mouth resuscitation. (S$_{22}$)	He pulled me to shore and started mouth-to-mouth resuscitation.	▪ Key word ▪ T.S.S ▪ Copy-verbatim
When I recovered, I opened my eyes to see my father crying.	When I rescued, I saw my father cry.	▪ Deletion ▪ Paraphrase ▪ Key word

186

Original sentence	Summary sentence	Summarizing strategy
(S_{23})		▪ T.S.S
My father was worried that the incident would scare me for life. (S_{26})	My father was so worried.	▪ Deletion ▪ Key word ▪ T.S.S
One day, he took me to the edge of the water and urged me to go in with him. (S_{27}) When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S_{30}) I plunged into the ocean and swam back to shore. (S_{33}) As my father proudly looked on, I knew I had overcome my fear. (S_{35})	And took me to the edge of water, my father challenged me to jump into the water and I jumped and my overcome my fear.	▪ Deletion ▪ Sentence combination ▪ Paraphrase ▪ Title word ▪ Location ▪ Key word ▪ T.S.S

Sample # 7

Original sentence	Summary sentence	Summarizing strategy
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S_{12})	I paddled and kicked hard, trying to remain above the surface.	▪ Deletion ▪ Title word ▪ T.S.S
Terror overcame me as I felt myself sinking to the bottom. (S_{14})	I felt myself sinking to the bottom.	▪ Deletion
Then I passed out. (S_{16})	Then I passed out.	▪ Location ▪ T.S.S ▪ Cope-verbatim
When my father emerged from the hut, he was horrified to see me struggling in the water some distance from the shore. (S_{17})	I struggling in the water some distance from the shore and my father was horrified to see me.	▪ Deletion ▪ Title word ▪ Location ▪ Key word ▪ T.S.S
He dived in and swam as hard as he could to the spot where I had gone under. (S_{18})	He dived in and swam.	▪ Deletion ▪ Key word ▪ T.S.S
He repeatedly dived under the water, frantically searching for my body. (S_{19})	He repeatedly dived under the water frantically searching for my body.	▪ Title word ▪ Cue ▪ T.S.S ▪ Cope-verbatim
He pulled me to shore and started mouth-to-mouth resuscitation. (S_{22})	Then my father pulled me to shore and started mouth-to-	▪ Key word ▪ T.S.S ▪ Cope-

187

Original sentence	Summary sentence	Summarizing strategy
mouth resuscitation.		- verbatim
When I recovered, I opened my eyes to see my father crying. (S23)	When I opened my eyes, I see my father crying.	- Deletion - Key word - T.S.S
In the days that followed, I was so traumatized that I would not go near the water. (S25)	In the days that followed, I was so traumatized that I would not go near the water.	- Title word - Location - Key word - T.S.S - Cope-verbatim
When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S30) I was overjoyed. (S34)	When I was confident enough, I was overjoyed.	- Deletion - Sentence combination - Title word - Key word - T.S.S - Copy-verbatim

Sample # 8

Original sentence	Summary sentence	Summarizing strategy
In one huge step, I stretched my arm as far as it could go and tried to grab the boat. (S9)	I stretched my arm as far as I could go to grab the boat.	- Deletion - Title word - Key word - T.S.S
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S12)	I paddled and kicked, trying to remain above the surface.	- Deletion - Title word - T.S.S
Terror overcame me as I felt myself sinking to the bottom. (S14)	Terror overcome me I felt myself sinking to the bottom.	- Cope-verbatim
When my father emerged from the hut, he was horrified to see me struggling in the water some distance from the shore. (S17)	My father horrified to see me struggling in the water from the shore.	- Deletion - Title word - Location - Key word - T.S.S
He grabbed it and yanked my head out of the water. (S21)	He grabbed it and yanked my head out of the water.	- Title word - T.S.S - Copy-verbatim
When I recovered, I opened my eyes to see my father crying. (S23)	I opened my eyes to see my father crying.	- Deletion - Key word - T.S.S
One day, he took me to the edge of the water and urged me to go in with him.	He took me to the edge of water and urged me to go	- Deletion - Title word - Key word

	in with him.	▪ T.S.S
(S27)		
In the following weeks, my father taught me how to swim. (S29)	My father taught me how to swim.	▪ Deletion ▪ Key word ▪ T.S.S
When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S30)	He challenged me to jump into the water.	▪ Deletion ▪ Title word ▪ Key word ▪ T.S.S

Sample # 9

Original sentence	Summary sentence	Summarizing strategy
I was enveloped in seawater. (S11) Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S12) However, the water was just too deep, the current was too strong and my body was too weak to fight. (S13)	I was enveloped in seawater and I panic-stricken the water is so deep.	▪ Deletion ▪ Sentence combination ▪ Paraphrase ▪ Title word ▪ T.S.S ▪ Copy-verbatim
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S12)	I paddled and kicked hard, trying to remain above the surface.	▪ Deletion ▪ Title word ▪ T.S.S
However, the water was just too deep, the current was too strong and my body was too weak to fight. (S13)	But the water too deep and the current too strong, my body so weak.	▪ Deletion ▪ Paraphrase ▪ Title word ▪ Cue ▪ T.S.S
Then I passed out. (S16)	Then I passed out.	▪ Location ▪ T.S.S ▪ Copy-verbatim
In the days that followed, I was so traumatized that I would not go near the water. (S25) In the following weeks, my father taught me how to swim. (S29)	I was so traumatized and would not go near the water, but my father taught me swim.	▪ Deletion ▪ Sentence combination ▪ Title word ▪ Location ▪ Key word ▪ T.S.S
When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S30)	I was confident enough and he took me at jetty.	▪ Deletion ▪ Location ▪ Key word ▪ T.S.S

189

Original sentence	Summary sentence	Summarizing strategy
However, the water was just too deep, the current was too strong and my body was too weak to fight. (S13) When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S30) I knew that the water was deep, but I was not looking at it. (S31) I plunged into the ocean and swam back to shore. (S33)	The water so deep but I jump into the water although I knew the water is deep but I plunged into the ocean and swam back to shore.	▪ Deletion ▪ Sentence combination ▪ Title word ▪ Cue ▪ Key word ▪ T.S.S
As my father proudly looked on, I knew I had overcome my fear. (S35)	I had overcome my fear.	▪ Deletion ▪ Location ▪ Key word ▪ T.S.S

Sample # 10

Original sentence	Summary sentence	Summarizing strategy
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S12)	I panic and fear.	▪ Deletion ▪ Paraphrase ▪ Title word ▪ T.S.S
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S12)	I paddled and kicked hard, trying to remain above the surface.	▪ Deletion ▪ Title word ▪ T.S.S
However, the water was just too deep, the current was too strong and my body was too weak to fight. (S13)	The water was just too deep and my body was weak to fight.	▪ Deletion ▪ Title word ▪ Cue ▪ T.S.S
Terror overcame me as I felt myself sinking to the bottom.14 Then his hand brushed against my hair. (S20) He grabbed it and yanked my head out of the water. (S21) He pulled me to shore and started mouth-to-mouth resuscitation. (S22)	I felt myself sinking to the bottom and my father save me.	▪ Deletion ▪ Sentence combination ▪ Key word ▪ T.S.S
In the days that followed, I was so traumatized that I	After that, I was so traumatized and not go near the water.	▪ Deletion ▪ Title word ▪ Location

Original sentence	Summary sentence	Summarizing strategy
would not go near the water. (S25)		▪ Key word ▪ T.S.S
My father was worried that the incident would scare me for life. (S26) In the following weeks, my father taught me how to swim. (S29)	My father worried that the incident wound scar me for life and he taught me how to swim.	▪ Deletion ▪ Sentence combination ▪ Key word ▪ T.S.S
When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S30) I was looking at my father, who showed great faith and confidence in me. (S32) I plunged into the ocean and swam back to shore. (S33)	I looking at my father and plunged into the water.	▪ Deletion ▪ Sentence combination ▪ Key word ▪ T.S.S
I was overjoyed. (S34)	I was overjoyed.	▪ Deletion ▪ Location ▪ Key word ▪ T.S.S
As my father proudly looked on, I knew I had overcome my fear. (S35)	My father proudly knew I overcame my fear.	▪ Deletion ▪ Location ▪ Key word ▪ T.S.S

Sample # 11

Original sentence	Summary sentence	Summarizing strategy
I was determined not to lose it. (S8)	I was determined not to lose it.	▪ Location ▪ T.S.S ▪ Copy-verbatim
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S12)	Panic-stricken, I paddled and kicked hard,trying to remain above the surface.	▪ Title word ▪ Key word ▪ T.S.S ▪ Copy-verbatim
He dived in and swam as hard as he could to the spot where I had gone under. (S18)	He dived in and swam as hard as he could to the spot where I had gone under.	▪ Title word ▪ T.S.S ▪ Copy-verbatim

Original sentence	Summary sentence	Summarizing strategy
He grabbed it and yanked my head out of the water. (S$_{21}$) He pulled me to shore and started mouth-to-mouth resuscitation. (S$_{22}$)	He grabbed it and yanked my head out of the water, and pulled me to shore and started mouth-to-mouth resuscitation.	▪ Sentence combination ▪ Title word ▪ Location ▪ Key word ▪ T.S.S ▪ Copy-verbatim
My father was worried that the incident would scare me for life. (S$_{26}$)	My father was worried that the incident.	▪ Deletion ▪ Key word ▪ T.S.S
In the following weeks, my father taught me how to swim. (S$_{29}$)	He taught me how to swim.	▪ Deletion ▪ Key word ▪ T.S.S
When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S$_{30}$)	He took me to a jetty and challenged me to jump into the water.	▪ Deletion ▪ Title word ▪ Key word ▪ T.S.S
I plunged into the ocean and swam back to shore.33	I plunged into the ocean and swam back to shore.	▪ Key word ▪ T.S.S ▪ Copy-verbatim
As my father proudly looked on, I knew I had overcome my fear. (S$_{35}$)	As my father proudly looked on I knew I had overcome my fear.	▪ Location ▪ Key word ▪ T.S.S ▪ Copy-verbatim

Sample # 12

Original sentence	Summary sentence	Summarizing strategy
One day, I was floating my boat on the water, carefully staying near the shore. (S$_1$)	One day, I was floating my boat.	▪ Deletion ▪ Title word ▪ Location ▪ Key word ▪ T.S.S
I started towards the shore as I saw my father turned away. (S$_4$) Unfortunately, I missed and fell into the ocean, not realizing how deep the water was and forgetting that I could not swim. (S$_{10}$)	When my father turns away, I fell into the ocean and forgetting I could not swim.	▪ Deletion ▪ Sentence combination ▪ Title word ▪ Key word ▪ T.S.S
Then I passed out. (S$_{16}$)	Then I passed out.	▪ Location ▪ T.S.S ▪ Copy-verbatim

Original sentence	Summary sentence	Summarizing strategy
When my father emerged from the hut, he was horrified to see me struggling in the water some distance from the shore. (S17) He dived in and swam as hard as he could to the spot where I had gone under. (S18) He repeatedly dived under the water, frantically searching for my body. (S19)	When my father emerged, he dived in and swam as hard to searching my body.	▪ Deletion ▪ Sentence combination ▪ Title word ▪ Location ▪ Key word ▪ Cue ▪ T.S.S
Then his hand brushed against my hair. (S20)	Then his hand brushed against my hair.	▪ Location ▪ T.S.S ▪ Copy-verbatim
He pulled me to shore and started mouth-to-mouth resuscitation. (S22)	He pulled and pulled me to shore and started mouth-to-mouth resuscitation.	▪ Key word ▪ T.S.S ▪ Copy-verbatim
When I recovered, I opened my eyes to see my father crying. (S23)	I opened my eyes to see my father crying.	▪ Deletion ▪ Key word ▪ T.S.S
My father was worried that the incident would scare me for life. (S26) In the following weeks, my father taught me how to swim. (S29)	My father worried the incident would scare me for life and taught me how to swim.	▪ Deletion ▪ Sentence combination ▪ Key word ▪ T.S.S

Sample # 13

Original sentence	Summary sentence	Summarizing strategy
In one huge step, I stretched my arm as far as it could go and tried to grab the boat. (S9) Unfortunately, I missed and fell into the ocean, not realizing how deep the water was and forgetting that I could not swim. (S10)	As I tried to grab the boat, I fall from the shore into the ocean.	▪ Deletion ▪ Sentence combination ▪ Title word ▪ Key word ▪ T.S.S
Unfortunately, I missed and fell into the ocean, not realizing how deep the water was and forgetting that I could not swim. (S10) However, the water was just too deep, the current was too strong and my body was too weak to fight. (S13)	The water was too deep and I could not swim.	▪ Deletion ▪ Sentence combination ▪ Title word ▪ Key word ▪ T.S.S

193

Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S_{12}) However, the water was just too deep, the current was too strong and my body was too weak to fight. (S_{13})	I try to save myself by trying to remain above the surface but it was too deep and the wave was too strong.	▪ Deletion ▪ Sentence combination ▪ Paraphrase ▪ Title word ▪ Cue ▪ T.S.S
Then I passed out. (S_{16})	Then I passed out.	▪ Location ▪ T.S.S ▪ Copy-verbatim
I was looking at my father, who showed great faith and confidence in me. (S_{32}) As my father proudly looked on, I knew I had overcome my fear. (S_{35})	I've overcome my fear by help from my father.	▪ Deletion ▪ Sentence combination ▪ Key word ▪ Location ▪ T.S.S
In the following weeks, my father taught me how to swim. (S_{29}) When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S_{30})	I learned to swim by him and accept his challenged with confident to jump into the water.	▪ Deletion ▪ Sentence combination ▪ Paraphrase ▪ Key word ▪ T.S.S
As my father proudly looked on, I knew I had overcome my fear. (S_{35})	With that, I had overcome my fear.	▪ Deletion ▪ Location ▪ Key word ▪ T.S.S

Sample # 14

Original sentence	Summary sentence	Summarizing strategy
But when I reached out to get my boat, a soft wind caught its tail and pushed it away from me. (S_{5})	The wind pushed it away from me.	▪ Deletion ▪ Title word ▪ Key word ▪ T.S.S
I took a couple of steps towards it, but the currents kept pushing the boat further and further away. (S_{6}) In one huge step, I stretched my arm as far as it could go and tried to grab the boat. (S_{9})	I took a huge step to grab it / stretched out my hand.	▪ Deletion ▪ Sentence combination ▪ Paraphrase ▪ Cue ▪ Location ▪ Key word ▪ T.S.S
Unfortunately, I missed and fell into the ocean, not realizing how deep the water was and forgetting that I could not swim. (S_{10})	I missed and fell into the ocean.	▪ Deletion ▪ Title word ▪ T.S.S

194

Unfortunately, I missed and fell into the ocean, not realizing how deep the water was and forgetting that I could not swim. (S10)	The water was deep and I could not swim.	▪ Deletion ▪ Title word ▪ T.S.S
Panic-stricken, I paddled and kicked hard, trying to remain above the surface. (S12)	I paddled hard, I kicked hard.	▪ Deletion ▪ Title word ▪ T.S.S
I gasped for air in desperation, but only salty water filled my throat and nostrils. (S15)	I gasped for air but filed them salty water.	▪ Deletion ▪ Generalization ▪ Title word ▪ T.S.S
Then I passed out. (S16)	I passed out.	▪ Location ▪ T.S.S ▪ Copy-verbatim
He dived in and swam as hard as he could to the spot where I had gone under. (S18) Then his hand brushed against my hair. (S20) He grabbed it and yanked my head out of the water. (S21) He pulled me to shore and started mouth-to-mouth resuscitation. (S22)	My father dived in end save me.	▪ Deletion ▪ Sentence combination ▪ Invention ▪ Key word ▪ T.S.S
In the days that followed, I was so traumatized that I would not go near the water. (S25)	I was traumatized and did not go near the water.	▪ Deletion ▪ Location ▪ Key word ▪ Title word T.S.S
In the following weeks, my father taught me how to swim. (S29)	My father taught me how to swim.	▪ Deletion ▪ Key word ▪ T.S.S
When I was confident enough, he took me to a jetty and challenged me to jump into the water. (S30)	My father challenges me to jump into the water from a jetty.	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S
I plunged into the ocean and swam back to shore. (S33)	I jumped in and swam back to the shore.	▪ Deletion ▪ Paraphrase ▪ Key word ▪ T.S.S
As my father proudly looked on, I knew I had overcome my fear. (S35)	I had overcome my fear.	▪ Deletion ▪ Location ▪ Key word ▪ T.S.S

$S_#$ − The position of the sentences in the original text.

Appendix D

Stop word List

Table D.1: Stop word List

able	kg	th
about	km	than
above	largely	thanx
abst	last	that
accordance	lately	that'll
according	later	thats
accordingly	latter	that've
across	latterly	the
actually	least	their
adj	less	theirs
after	lest	them
afterwards	let	themselves
again	lets	then
against	likely	thence
ah	line	there
all	little	thereafter
almost	'll	thereby
alone	ltd	thered
along	mainly	therein
already	many	there'll
also	may	thereof
although	maybe	therere
always	me	theres
am	mean	thereto
among	means	thereupon
amongst	meantime	there've
an	meanwhile	these
and	merely	they
announce	mg	theyd
another	might	they'll
any	million	theyre
anybody	miss	they've
anyhow	ml	this
anymore	more	those
anyone	moreover	thou
anything	most	though
anyway	mostly	thoughh
anyways	mr	thousand
anywhere	mrs	throug
apparently	much	through
approximately	mug	throughout
are	must	thru
aren	my	thus
arent	myself	til

arise	na	tip
around	name	to
as	namely	together
aside	nay	too
at	nd	toward
auth	near	towards
available	nearly	tried
away	necessarily	tries
awfully	necessary	truly
be	neither	try
because	never	trying
been	nevertheless	ts
before	new	twice
beforehand	next	two
begin	nine	un
beginning	ninety	under
beginnings	no	unfortunately
begins	nobody	unless
behind	non	unlike
being	none	unlikely
believe	nonetheless	until
below	noone	unto
beside	nor	up
besides	normally	upon
between	nos	ups
beyond	not	us
biol	noted	usefully
both	nothing	usefulness
brief	now	usually
briefly	nowhere	value
but	obtain	various
by	obtained	've
ca	obviously	very
came	of	via
can	off	viz
cannot	often	vol
can't	oh	vols
cause	ok	vs
causes	okay	w
certain	old	want
co	on	wants
com	once	was
could	one	wasn't
couldnt	ones	way
date	only	we
did	onto	wed
didn't	or	welcome
different	ord	we'll
do	other	went
does	others	were
doesn't	otherwise	weren't
doing	ought	we've

done	our	what
don't	ours	whatever
down	ourselves	what'll
downwards	out	whats
due	outside	when
during	over	whence
each	overall	whenever
ed	owing	where
edu	own	whereafter
effect	page	whereas
eg	pages	whereby
eight	part	wherein
eighty	particular	wheres
either	particularly	whereupon
else	past	wherever
elsewhere	per	whether
ending	perhaps	which
enough	placed	while
especially	please	whim
et	plus	whither
et-al	poorly	who
etc	possible	whod
even	possibly	whoever
ever	potentially	whole
every	pp	who'll
everybody	predominantly	whom
everyone	present	whomever
everything	previously	whos
everywhere	primarily	whose
ex	probably	why
except	promptly	widely
far	proud	willing
few	provides	wish
ff	que	with
fifth	quickly	within
first	quite	without
five	qv	won't
fix	ran	words
for	rather	world
former	rd	would
formerly	re	wouldn't
forth	readily	www
found	really	yes
four	recent	yet
from	recently	you
further	ref	youd
furthermore	refs	you'll
had	regarding	your
hardly	regardless	youre
has	regards	yours
hasn't	related	yourself
have	relatively	yourselves

haven't	research	you've
having	respectively	zero
he	resulted	
hed	resulting	specified
hence	results	specify
her	right	specifying
here	same	still
hereafter	says	stop
hereby	sec	strongly
herein	section	sub
heres	seem	substantially
hereupon	seemed	successfully
hers	seeming	such
herself	seems	sufficiently
hes	self	suggest
hi	selves	sup
hid	sent	sure
him	seven	
himself	several	it'll
his	shall	its
hither	she	itself
home	shed	i've
how	she'll	just
howbeit	shes	specifically
however	should	
hundred	shouldn't	itd
id	showns	
ie	significant	it
if	significantly	
i'll	similar	isn't
im	similarly	
immediate	since	soon
immediately	six	
importance	slightly	
important	so	
in	some	
inc	somebody	
indeed	somehow	
index	someone	
information	somethan	
instead	something	
into	sometime	
invention	sometimes	
inward	somewhat	
is	somewhere	

APPENDIX E

ANALYSIS

Analysis on Summary Sentences

- **Some samples**

Sample # 1

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
The hypnotist tries to obtain his subject's co-operation by pointing out to him the advantages to be secured by the hypnosis such as the help in curing a nervous illness to be derived from patients.	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S
It is sometimes helpful to concentrate subject's attention on some bright object dangled just above eye-level, thus forcing him to look slightly upwards.	▪ Copy-verbatim ▪ T.S.S ▪ Key word ▪ Cue	▪ T.S.S ▪ Key word ▪ Cue ▪ Deletion
Successful suggestions of this kind are instrumental in deepening the hypnotic trance.	▪ Deletion ▪ Key word ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Location ▪ T.S.S
He will take up any suggestion hypnotist puts forward and act on it to best of his ability.	▪ Key word ▪ T.S.S ▪ Copy-verbatim	▪ Key word ▪ Deletion ▪ T.S.S

Sample # 2

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
The hypnotist tries to obtain his subject's co-operation by pointing out to him the advantages to be secured by the hypnosis.	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ T.S.S
But there is a common that have been used by hypnotists that is hypnotist tried to obtain his subject's co-operation by pointing-out to him the advantages to be secured by the hypnosis.	▪ Sentence combination ▪ Paraphrase ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S	▪ Sentence combination ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S
The most obvious type of phenomenon can be elicited, that is a tremendous increase in the subject's suggestibility.	▪ Sentence combination ▪ Deletion ▪ Key word ▪ Cue ▪ Location ▪ T.S.S	▪ Sentence combination ▪ Deletion ▪ Key word ▪ Cue ▪ Location ▪ T.S.S
This tremendous increase suggestibility is exploited on the stage to induce people to do foolish and ridiculous act.	▪ Copy-verbatim ▪ T.S.S ▪ Key word	▪ Deletion ▪ T.S.S ▪ Key word
Such practice are not to be encouraged because they go counter to the ideal of human dignity and are not a kind of way which hypnosis ought to be used.	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S

Sample # 3

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
There are many methods of producing hypnosis.	▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S ▪ Deletion	▪ Key word ▪ Title word ▪ Location ▪ T.S.S ▪ Deletion ▪ Cue

201

The subject is asked to lie down and external stimulation is reduced to a minimum as far as possible.	- Deletion - Sentence combination - Key word - Location - T.S.S	- Deletion - Sentence combination - Key word - Location - T.S.S
Some small object will be dangled above eye-level of the subject, and then his eyes will be closing.	- Deletion - Sentence combination - Key word - Cue - T.S.S	- Deletion - Sentence combination - Key word - Cue - T.S.S
It would not be true to say, however, all suggestions are accepted, but in very deepest trance.	- Copy-verbatim - Paraphrase - Key word - Cue - Location - T.S.S	- Deletion - Key word - Cue - Location - T.S.S
A well-known story may be quoted to illustrate.	Copy-verbatim	Deletion
He, the subject will hear to the hypnotist voice.	- Deletion - Key word - T.S.S	- Deletion - Key word - T.S.S
Then, the subject will take up any suggestion the hypnotist puts and act on his best ability.	- Deletion - Key word - T.S.S	- Deletion - Key word - T.S.S

Sample # 4

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
The hypnotist uses different ways for hypnosis. The hypnotist should secure the information collected from their patients.	- Deletion - Sentence combination - Key word - Title word - Location - T.S.S - Invention	- Deletion - Sentence combination - Key word - Title word - Location - T.S.S
They are told on the consequences of hypnosis and asked to lie down on a couch or sit in an easy chair.	- Deletion - Sentence combination - Key word - Title word - Location - T.S.S	- Deletion - Sentence combination - Key word - Title word - Location - T.S.S

External stimulation is reduced to minimum and their attention is concentrated on a small bright object right above eye level leading to fatigue of eye-muscles.	- Deletion - Sentence combination - Key word - Location - Cue - T.S.S	- Deletion - Sentence combination - Key word - Location - Cue - T.S.S
Subject is asked to lie down on a couch, or sit in an easy-chair. It is sometimes helpful to concentrate the subject's attention.	- Deletion - Sentence combination - Copy-verbatim - Key word - Location - Cue - T.S.S	- Deletion - Sentence combination - Key word - Location - Cue - T.S.S
A well-known story may be quoted to illustrate this.	Copy-verbatim	Copy-verbatim

Sample # 5

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
Hypnosis produced by many methods.	- Deletion - Key word - Title word - Location - Cue - T.S.S	- Deletion - Key word - Title word - Location - Cue - T.S.S
The hypnotist tries to obtain his subject's co-operation by pointing out to him the advantages to be secured, such as the help in curing a nervous illness.	- Deletion - Key word - Title word - T.S.S	- Deletion - Key word - Title word - T.S.S
The patient is reassured about any possible dangers he might suspect to be present in hypnosis.	- Deletion - Key word - Title word - Location - T.S.S	- Deletion - Key word - Title word - Location - T.S.S
External stimulation sometimes helpful to concentrate the subject's attention.	- Deletion - Key word - Location - Cue - Sentence combination	- Deletion - Key word - Location - Cue - Sentence combination

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	RDSSIA
	T.S.S	T.S.S
This leads to a fatigue of the eye-muscles and feeling tired so that his eyes are closing, then the hypnotist starts to talk to the subject.	- Deletion - Key word - Cue - Sentence combination - T.S.S	- Deletion - Key word - Cue - Sentence combination - T.S.S
In a susceptible subject, a light trance is induced and the hypnotist now begins deepen.	- Deletion - Key word - Cue - T.S.S	- Deletion - Key word - Cue - T.S.S
She immediately awakened her trance, slapped his face, and flounced out room, much to his discomfiture.	- Copy-verbatim - Key word - Location - T.S.S	- Deletion - Key word - Location - T.S.S

Sample # 6

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	RDSSIA
The hypnotist tries to obtain subject's co-operation.	- Deletion - Key word - Title word - T.S.S	- Deletion - Key word - Title word - T.S.S
Next, subject is asked to lie down on a couch, or sit in an easy-chair.	- Deletion - Key word - Location - T.S.S	- Deletion - Key word - Location - T.S.S
Hypnotist begins talk to the subject in a soft tone of voice.	- Deletion - Key word - T.S.S	- Deletion - Key word - T.S.S
Hypnotist now begins to deepen this trance and to test the reactions of the subject.	- Deletion - Key word - Cue - T.S.S	- Deletion - Key word - Cue - T.S.S
Subject will take up any suggestions the hypnotist puts forward.	- Deletion - Key word - T.S.S	- Deletion - Key word - T.S.S

Subject will go counter to the ideal of human dignity.	■ Deletion ■ Key word ■ Title word ■ Location ■ T.S.S	■ Deletion ■ Key word ■ Title word ■ Location ■ Cue ■ T.S.S
The subjects should sit in a quiet place, without any noise and then, look to a point.	■ Deletion ■ Key word ■ Location ■ Cue ■ T.S.S ■ Sentence combination ■ Invention	■ Deletion ■ Key word ■ Location ■ Cue ■ T.S.S ■ Sentence combination

Sample # 7

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
A well-known story may quoted to illustrate.	■ Copy-verbatim	■ Deletion
Couch or an easy-chair is usually provided for the subject to lie down.	■ Deletion ■ Key word ■ Location ■ T.S.S	■ Deletion ■ Key word ■ Location ■ T.S.S
She had been hypnotized deeply he called away, and handed over demonstration of his assistants.	■ Key word ■ T.S.S ■ Copy-verbatim	■ Deletion ■ Key word ■ T.S.S
Hypnotist sometimes used some small bright object dangled just above eye-level.	■ Deletion ■ Key word ■ Cue ■ T.S.S	■ Deletion ■ Key word ■ Cue ■ T.S.S
Then, he will talk in a soft tone of voice, giving suggestions so that the subject cannot hear anything but the hypnotist's voice.	■ Deletion ■ Key word ■ T.S.S	■ Deletion ■ Key word ■ T.S.S
The most obvious one is a tremendous increase in the subject's suggestibility.	■ Deletion ■ Key word ■ Location ■ T.S.S	■ Deletion ■ Key word ■ Location ■ Cue ■ T.S.S

Summary sentences	Summarizing Strategies / Methods Identified	
	Deletion **Key word** **Title word** **Sentence combination** **Location** **Cue** **T.S.S**	**Deletion** **Key word** **Title word** **Sentence combination** **Location** **Cue** **T.S.S**
But the suggestions should not counter the ideal of human dignity or ethical and moral conceptions held by the subject.	▪ Deletion ▪ Key word ▪ Title word ▪ Sentence combination ▪ Location ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ Sentence combination ▪ Location ▪ Cue ▪ T.S.S

Sample # 8

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
The hypnotist tries to obtain his subject's co-operation by pointing out to him the advantages to be secured by the hypnosis.	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S
The patient is reassured about any possible dangers he might suspect to be present in hypnosis, and he may also be told that is not a sign instability or weakness.	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S
External stimulation is reduced to a minimum by drawing the curtains and excluding.	▪ Deletion ▪ Location ▪ T.S.S	▪ Deletion ▪ Location ▪ T.S.S
Successful suggestions of this kind are instrumental in deepening the hypnosis trance until final.	▪ Deletion ▪ Key word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Location ▪ T.S.S
This is particularly true when a suggestion is made which is contrary to the ethical and moral conceptions held.	▪ Deletion ▪ Key word ▪ T.S.S	▪ Deletion ▪ Key word ▪ T.S.S

Sample # 9

Summary sentences	Summarizing Strategies / Methods Identified

	Human expert	RDSSIA
Hypnosis will try to obtain his subject's co-operation by pointing out to him the advantages to be secured.	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S
Next, the subject is asked to lie down on a couch or sit in an easy-chair and some of it helpful to concentrate the subject attention on some small bright object dangled just above eye-level.	▪ Deletion ▪ Key word ▪ Sentence combination ▪ Location ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Sentence combination ▪ Location ▪ Cue ▪ T.S.S
Besides that, it will take up any suggestion the hypnotist puts up and act in the best of their ability.	▪ Deletion ▪ Key word ▪ T.S.S	▪ Deletion ▪ Key word ▪ T.S.S
Most common method is something along these lines.	▪ Copy-verbatim	▪ Deletion ▪ Location ▪ T.S.S

Sample # 10

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	RDSSIA
Hypnosis is a means of communication between the conscious mind and the subconscious mind.	▪ Invention ▪ Title word	▪ Deletion ▪ Key word ▪ Location ▪ Title word ▪ Cue ▪ T.S.S
For instance, the help in curing a nervous illness to be derived from the patient's remembering in the trance certain events which otherwise are inaccessible to his memory.	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S
The hypnotists begin in a soft tone of voice with subject.	▪ Deletion ▪ Key word ▪ T.S.S	▪ Deletion ▪ Key word ▪ T.S.S
Successful suggestions of kind are instrumental in deepening hypnotic trance, finally, in particularly good subjects; all phenomena which will be	▪ Copy-verbatim ▪ Key word ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Location ▪ T.S.S

Summary sentences	Summarizing Strategies / Methods Identified	
discussed presently can be elicited.		
Young man, lacking the seriousness of purpose.	▪ Deletion ▪ Key word ▪ T.S.S	▪ Deletion ▪ Key word ▪ T.S.S

Sample # 11

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
Hypnosis can be producing by many methods.	▪ Deletion ▪ Key word ▪ Title word ▪ Location ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ Location ▪ Cue ▪ T.S.S
First, hypnotist tries to obtain his subject's attention by pointing out.	▪ Deletion ▪ Key word ▪ Sentence combination ▪ Title word ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Sentence combination ▪ Title word ▪ Cue ▪ T.S.S
Hypnosis also can help in curing a nervous illness or remembering.	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S
Next method of hypnosis is the subject is asked to lie down on a couch or sit in an easy-chair.	▪ Deletion ▪ Key word ▪ Sentence combination ▪ Title word ▪ Cue ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Sentence combination ▪ Title word ▪ Cue ▪ Location ▪ T.S.S
Disruptive noises is something helpful to concentrate the subject attention.	▪ Deletion ▪ Key word ▪ Sentence combination ▪ Cue ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Sentence combination ▪ Cue ▪ Location ▪ T.S.S

Summary sentences	Human expert	RDSSIA
This tremendous increase in suggestibility is exploited on stage to induce people to do foolish and ridiculous acts.	▪ Copy-verbatim	▪ Deletion ▪ Key word ▪ T.S.S

Sample # 12

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
The patient is reassured about any possible dangers he might suspect to be present in hypnosis.	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S
This leads quickly to a fatigue of the eye-muscles and thus facilitates his acceptance of the suggestion that he is feeling tired and that his eyes are closing.	▪ Copy-verbatim ▪ Key word ▪ Cue ▪ T.S.S	▪ Copy-verbatim ▪ Key word ▪ Cue ▪ T.S.S
The hypnotist now begins to talk to the subject in a soft tone of voice.	▪ Deletion	▪ Deletion ▪ Key word ▪ T.S.S
The responsible in large measure for all the others, is a tremendous increase in subject's suggestibility.	▪ Deletion ▪ Key word ▪ Location ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Location ▪ Cue ▪ T.S.S
Suggestion the hypnotist puts forward and act on it to the best of ability.	▪ Deletion ▪ Key word ▪ T.S.S	▪ Deletion ▪ Key word ▪ T.S.S

Sample # 13

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
There are many methods of producing hypnosis.	▪ Deletion ▪ Key word ▪ Title word ▪ Cue	▪ Deletion ▪ Key word ▪ Title word ▪ Cue

Summary sentences	Human expert	RDSSIA
	▪ Location ▪ T.S.S	▪ Location ▪ T.S.S
Firstly, the hypnotist must try to obtain a subject's co-operation by pointing out to him the advantages to be secured by hypnosis.	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S
Secondly, hypnosis can be produced by reducing external stimulation to a minimum by drawing the curtains and excluding.	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S ▪ Sentence combination	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S ▪ Sentence combination
Besides, hypnotist can give suggestions which are more and more difficult of execution in order to test the reaction of the subject.	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S
Additionally, a suggestion made by hypnotists must be contrary to the ethical and moral conceptions held by the subject.	▪ Deletion ▪ Key word ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Location ▪ T.S.S
External stimulation is decreased to a minimum by drawing the curtains and excluding, all disruptive noises.	▪ Copy-verbatim ▪ T.S.S	▪ Deletion ▪ Location ▪ T.S.S ▪ paraphrase

Sample # 14

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
The hypnotist tries to obtain his subject co-operation and the patient is reassured about any possible danger he might suspect to present in hypnosis.	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S

Sometimes hypnotist is helpful to concentrate the subject's attentions but leads quickly to a fatigue of the eye-muscles, thus facilitate his acceptance of the suggestion.	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S ▪ Sentence combination	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S ▪ Sentence combination
A tremendous increase in the subject's suggestibility that may be responsible in large measure for all the others.	▪ Deletion ▪ Key word ▪ Cue ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Cue ▪ Location ▪ T.S.S
Pointing the subject to be something or someone are not encouraged because they go counter to the ideal of human dignity and not the way hypnosis ought to be used.	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S

Sample # 15

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	RDSSIA
The hypnotist is to obtain by pointing out if him, to be secured.	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S
Firstly, they will ask to lie down on couch.	▪ Deletion ▪ Key word ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Location ▪ T.S.S
Secondly, it is helpful to concentrate the subjects attention on some small bright object dangled just above eye-level.	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S
Next, try to talk to subject in a soft tone of voice.	▪ Deletion ▪ Key word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S

211

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	RDSSIA
In addition, he will ask top clap his hand together and tells it is impossible to separate his hands.	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S
Last but not least, exploited on stage to induce people to do foolish and ridiculous acts and not to be encouraged.	▪ Deletion ▪ Key word ▪ T.S.S	▪ Deletion ▪ Key word ▪ T.S.S

Sample # 16

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	RDSSIA
There are many method of producing hypnosis.	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S
Some method for hypnosis is the hypnotist tries to obtain his subject's co-operation by pointing out to him the advantages to be secured by the hypnosis, such as, the help in curing a nervous illness to be derived from patient's remembering in the trance certain events which otherwise are inaccessible to his memory.	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S ▪ Sentence combination	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S ▪ Sentence combination
Tremendous increase in the subject's suggestibility.	▪ Deletion ▪ Cue ▪ Location ▪ T.S.S	▪ Deletion ▪ Cue ▪ Location ▪ T.S.S
He will take up any suggestion the hypnotist puts forward and act on it to the best of his ability.	▪ Copy-verbatim ▪ T.S.S	▪ Copy-verbatim ▪ Key word ▪ T.S.S

Such practices are not be encouraged because they go counter to the ideal of human dignity and are not the kind of way in which hypnosis ought to be used.	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S

Sample # 17

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	**RDSSIA**
The methods of producing hypnosis are to obtain his subject's co-operation by pointing out the advantages to be secured.	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S ▪ Sentence combination	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S ▪ Sentence combination
Besides, to help in curing a nervous illness.	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ T.S.S
Next, the subject asked to lie down on a couch.	▪ Deletion ▪ Key word ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Location ▪ T.S.S
External stimulation is helpful to concentrate the subject's attention.	▪ Deletion ▪ Key word ▪ Cue ▪ Location ▪ T.S.S ▪ Sentence combination	▪ Deletion ▪ Key word ▪ Cue ▪ Location ▪ T.S.S ▪ Sentence combination
This leads quickly to a fatigue.	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S
The hypnotist talk to subject in a soft tone of voice.	▪ Deletion ▪ Key word ▪ T.S.S	▪ Deletion ▪ Key word ▪ T.S.S
A light trance is thus induced after a few minutes.	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S

Besides, it'll ask the subject to claps his hands together.	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S
The type of phenomena is tremendous which increase in the subject's suggestibility.	▪ Deletion ▪ Key word ▪ Cue ▪ Location ▪ T.S.S ▪ Sentence combination	▪ Deletion ▪ Key word ▪ Cue ▪ Location ▪ T.S.S ▪ Sentence combination
Practices are not to be encouraged as they go counter to the ideal of human dignity.	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S	▪ Deletion ▪ Key word ▪ Title word ▪ Cue ▪ Location ▪ T.S.S

Sample # 18

Summary sentences	Summarizing Strategies / Methods Identified	
	Human expert	RDSSIA
External stimulation is reduced to a minimum by drawing the curtains and extending, all disruptive noises.	▪ Deletion ▪ Location ▪ T.S.S	▪ Deletion ▪ Location ▪ T.S.S
The hypnotist now begins to talk to the subject in a soft tone of voice, repeating suggestions to the effect that the subject is getting tired.	▪ Deletion ▪ Key word ▪ T.S.S	▪ Deletion ▪ Key word ▪ T.S.S
The hypnotist now begins to deepen this trance and to test the reactions of the subject by giving suggestions which are more and more difficult of execution.	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S	▪ Deletion ▪ Key word ▪ Cue ▪ T.S.S
Subject is wanted to lie down on a couch, or sit in easy-chair.	▪ Copy-verbatim ▪ Key word ▪ T.S.S ▪ Paraphrase	▪ Deletion ▪ Key word ▪ Location ▪ T.S.S
This tremendous increase is suggestibility often exploited on the stage to induce people to do foolish and ridiculous acts.	▪ Deletion ▪ Key word ▪ T.S.S	▪ Deletion ▪ Key word ▪ T.S.S

Table E.1 Summarizing strategies identified by RDSSIA

Summary	No. of summary sentences	No. of summarizing strategies					No. of methods			
		D	S.C	P	TSS	CP	Cue	Title	Location	Key word
1	4	4	0	0	4	0	1	1	1	4
2	5	4	2	0	5	0	4	2	3	5
3	6	7	2	0	6	0	3	1	2	6
4	4	4	4	0	4	1	2	2	3	4
5	7	7	2	0	7	0	4	3	4	6
6	7	7	1	0	7	0	3	2	3	7
7	6	7	1	0	6	0	2	1	3	6
8	5	5	0	0	5	0	0	2	2	4
9	4	4	1	0	4	0	0	1	2	3
10	5	5	0	0	4	0	1	2	2	5
11	6	5	3	0	6	0	4	4	3	6
12	5	4	0	0	5	1	2	1	1	5
13	6	5	1	1	6	0	3	3	3	5
14	4	4	1	0	4	0	3	2	2	4
15	6	6	0	0	6	0	3	1	1	5
16	5	4	1	0	5	1	4	3	4	4
17	10	10	3	0	10	0	7	3	5	10
18	4	5	0	0	4	0	1	0	1	4
19	4	4	2	0	4	0	2	1	2	4
20	4	3	2	0	4	1	2	2	2	4

Where,

D: Deletion,   **SC**: Sentence combination,   **G**: Generalization,   **P**: Paraphrase,

TSS: Topic Sentence Selection,   **Inv**: Invention,   **CV**: Copy –verbatim.

Table E.2 Summarizing strategies identified by RDSSIA

Summary	No. of summary sentences	No. of summarizing strategies					No. of methods			
		D	S.C	P	TSS	CP	Cue	Title	Location	Key word
21	5	5	1	0	5	0	3	1	3	5
22	6	4	1	0	6	2	4	3	3	6
23	6	6	0	0	6	0	3	3	3	5
24	5	4	0	0	5	0	0	1	1	4
25	6	7	0	0	6	0	2	1	2	6
26	6	5	0	0	6	1	1	0	2	5
27	8	5	1	0	8	2	3	1	2	8
28	11	9	2	0	11	2	3	3	7	11
29	9	5	1	0	9	1	2	2	4	8
30	7	6	2	0	7	1	2	2	4	6
31	9	9	0	1	9	0	3	3	5	8
32	7	7	1	0	7	0	3	2	3	7
33	8	8	2	0	8	0	3	1	4	7
34	4	4	1	0	4	0	1	1	3	4
35	5	4	0	0	5	1	2	2	3	4
36	4	4	0	0	4	0	2	1	1	4
37	5	5	2	0	5	0	2	3	3	5
38	2	2	1	0	2	0	1	1	2	2
39	6	5	1	0	6	0	3	3	5	5
40	4	4	1		4		1	2	2	4

Where,

D: Deletion,   **SC**: Sentence combination,   **G**: Generalization,   **P**: Paraphrase,

TSS: Topic Sentence Selection,   **Inv**: Invention,   **CV**: Copy –verbatim.

Table E.3 Summarizing strategies identified by RDSSIA

Summary	No. of summary sentences	No. of summarizing strategies					No. of methods			
		D	S.C	P	TSS	CP	Cue	Title	Location	Key word
41	5	3	0	0	4	1	1	1	3	4
42	6	5	0	0	6	1	1	1	2	6
43	8	7	0	1	8	1	4	1	4	7
44	7	6	1	0	7	1	3	3	4	7
45	10	9	2	1	10	1	1	1	4	10
46	7	6	0	0	7	1	1	3	5	7
47	11	8	0	0	11	2	4	1	3	9
48	5	3	0	0	5	2	1	2	1	5
49	6	6	0	0	6	0	2	0	2	5
50	8	7	0	1	8	1	2	2	5	4
51	7	6	1	0	7	1	4	2	3	7
52	4	2	0	0	4	0	1	1	2	4
53	11	7	1	0	11	4	5	0	6	10
54	9	9	2	0	9	1	5	2	4	9
55	3	3	0	0	3	0	2	1	3	3
56	8	8	0	1	8	0	4	2	1	7
57	8	5	1	0	7	2	5	1	4	8
58	4	4	0	0	4	0	1	0	1	4

Where,

D: Deletion,   **SC**: Sentence combination,   **G**: Generalization,   **P**: Paraphrase,

TSS: Topic Sentence Selection,   **Inv**: Invention,   **CV**: Copy –verbatim.

Table E.4 Precision, Recall and F-score

Summary	A	B	C	Precision	Recall	F-score
1	4	2	2	0.67	0.67	0.67
2	5	0	2	1.00	0.71	0.83
3	6	0	2	1.00	0.75	0.86
4	5	0	2	1.00	0.71	0.83
5	7	1	1	0.88	0.88	0.88
6	7	1	1	0.88	0.88	0.88
7	7	2	2	0.78	0.78	0.78
8	5	1	0	0.83	1.00	0.91
9	3	0	0	1.00	1.00	1.00
10	5	6	2	0.45	0.71	0.56
11	5	0	0	1.00	1.00	1.00
12	5	0	2	1.00	0.71	0.83
13	6	3	1	0.67	0.86	0.75
14	4	0	0	1.00	1.00	1.00
15	6	1	0	0.86	1.00	0.92
16	5	1	0	0.83	1.00	0.91
17	10	0	0	1.00	1.00	1.00
18	5	2	2	0.71	0.71	0.71
19	3	2	1	0.60	0.75	0.67
20	4	0	2	1.00	0.67	0.80
21	6	1	1	0.86	0.86	0.86
22	6	2	3	0.75	0.67	0.71
23	6	4	1	0.60	0.86	0.71
24	5	0	3	1.00	0.63	0.77
25	7	0	5	1.00	0.58	0.74
26	6	1	1	0.86	0.86	0.86
27	8	1	3	0.89	0.73	0.80
28	11	1	2	0.92	0.85	0.88
29	9	1	4	0.90	0.69	0.78
30	7	1	1	0.88	0.88	0.88
31	9	2	2	0.82	0.82	0.82
32	7	2	1	0.78	0.88	0.82

33	8	2	1	0.80	0.89	0.84
34	4	1	1	0.80	0.80	0.80
35	5	2	0	0.71	1.00	0.83
36	3	1	1	0.75	0.75	0.75
37	5	0	0	1.00	1.00	1.00
38	2	0	0	1.00	1.00	1.00
39	6	0	0	1.00	1.00	1.00
40	4	0	0	1.00	1.00	1.00
41	5	0	2	1.00	0.71	0.83
42	6	0	4	1.00	0.60	0.75
43	8	2	1	0.80	0.89	0.84
44	7	0	0	1.00	1.00	1.00
45	10	1	3	0.91	0.77	0.83
46	7	0	1	1.00	0.88	0.93
47	11	0	0	1.00	1.00	1.00
48	5	0	0	1.00	1.00	1.00
49	6	1	1	0.86	0.86	0.86
50	8	2	3	0.80	0.73	0.76
51	7	0	1	1.00	0.88	0.93
52	4	1	0	0.80	1.00	0.89
53	11	2	4	0.85	0.73	0.79
54	9	2	3	0.82	0.75	0.78
55	3	1	3	0.75	0.50	0.60
56	8	1	1	0.89	0.89	0.89
57	7	2	1	0.78	0.88	0.82
58	4	1	1	0.80	0.80	0.80

Table E.5: No. of the same summarizing strategies identified by RDSSIA and Human expert

Summary	Number of summary sentences which have same summarizing strategies	Number of summary sentences in each summary
1	2	4
2	3	5
3	5	7
4	3	5
5	7	7
6	5	7
7	4	7
8	4	5
9	3	4
10	3	5
11	5	6
12	4	5
13	5	6
14	4	4
15	5	6
16	4	5
17	10	10
18	4	5
19	2	4
20	3	4
21	5	6
22	3	6
23	5	6
24	4	5
25	6	7
26	5	6
27	7	8
28	10	11
29	7	9
30	6	7

31	8	9
32	5	7
33	6	8
34	3	4
35	4	5
36	2	4
37	5	5
38	2	2
39	6	6
40	4	4
41	4	5
42	5	6
43	6	8
44	7	7
45	9	10
46	6	7
47	11	11
48	5	5
49	5	6
50	6	8
51	6	7
52	3	4
53	10	11
54	8	10
55	2	3
56	7	8
57	6	7
58	3	4