

# An Algorithm for Summarizing Strategies Identification

Norisma IDRIS, Mohd Sapiyan BABA, Rukaini ABDULLAH

Artificial Intelligence Dept, Faculty of Computer Science & IT, University Malaya,  
50603 Kuala Lumpur, Malaysia  
[norisma@um.edu.my](mailto:norisma@um.edu.my)

**Abstract:** Summarizing is a process of identifying important information from a text. Experts employ several important strategies to produce good summaries. Unfortunately, students do not acquire the summarization skills. Thus, this paper proposed an algorithm to identify the summarizing strategies employed by students using summary sentence decomposition. The strategies used by experts are identified and translated into a set of heuristic rules where the algorithm is developed based on the heuristic rules.

**Keywords:** summarizing strategies, summary sentence decomposition, heuristic rules

## 1. Introduction

Summary writing is an important skill which involves multiple cognitive activities such as understanding a text and identifying relevant content to generate a summary. The difference between summary writing and other writings such as essay writing is that the production of a summary dependent on the existing text [1]. In Malaysia, summary writing has been part of the English Language syllabus for many years [2].

Summarization is one of the best learning techniques to evaluate student's comprehension [3]. Hence, automated summarization assessment has drawn a lot of interest in recent years, e.g. *Summary Street*® [4] and *Laburpen Ebaluaka Automatikoa (LEA)* [3]. *Summary Street*® provides an environment where students can get feedback about the content knowledge, writing mechanics, length, redundancy and plagiarism. *LEA* is an automatic summary evaluation environment to train students in summarization skills. These two tools employed *Latent Semantic Analysis (LSA)* to construct the semantic representations that mirror the way human knowledge is structured. Previous study has shown that student's difficulties in summarizing were linked to students' use of strategic summarization skills [5]. In addition, previous study on students' summaries suggested that students' performances do not reflect directly their skills in summarizing [6]. Thus, rather than evaluating the content of the summary, which has been carried out by many researchers, we study on how to identify the summarizing strategies used by students.

## 2. Designing Heuristic Rules based on Experts' Summarizing Strategies

The design of the heuristic rules is based on the expert summarizing skills which are acquired by studying the experts' summaries. The study was conducted to identify the experts' strategies and how the strategies are used to produce the summary sentences. Based on the basic rules proposed by Brown and Day [7], we identified 8 types of strategies that are commonly used by the experts. These strategies are deletion, sentence combination, topic sentence selection, paraphrase, generalization, syntactic transformation, sentence reordering and invention. We used position-based-method to analyze the summary sentence. Thus, to represent the sentences and words in a text, we use this notation for our discussion. If  $T$  is a text consisting of  $m$  sentences,  $t_i$  the  $i^{th}$  sentence, then,  $T = \{ t_i \}; i = 1, 2, 3, \dots, m$ . Hence, for sentence  $t_i$  comprising a string of  $n_i$  words,  $t_{ij}$ , then,  $t_i = \{ t_{ij} \}; j = 1, 2, 3, \dots, n_i$ . Similarly for summary text,  $S$ , where every summary sentence,  $s_i$  comprises a string of words, represented as,  $s_i = \{ s_{ij} \}; i = 1, 2, 3, \dots, m$  and  $j = 1, 2, 3, \dots, n_i$ . The experts' strategies were translated into a set of heuristic rules as shown in Table 1 below:

Table 1. A set of heuristic rules for detecting summarizing strategies

Rule	Heuristic Rules
Deletion	<p><math>s_i</math> is produced by deletion if:</p> <ul style="list-style-type: none"> <li>the words in <math>s_i</math> are found in the same sentence in <math>T</math>, say <math>t_j</math></li> <li>the words in <math>s_i</math> are located near to each other in <math>T</math></li> <li>the number of words in <math>s_i</math> are less than <math>t_j</math></li> </ul>
Sentence Combination	<p><math>s_i</math> is produced by sentence combination if:</p> <ul style="list-style-type: none"> <li>the words in <math>s_i</math> are found in different sentences in <math>T</math></li> <li>the combined sentences are located near to each other in <math>T</math></li> </ul>
Syntactic Transformation	<p><math>s_i</math> is produced by syntactic transformation if:</p> <ul style="list-style-type: none"> <li>the words in <math>s_i</math> are found in the same sentence in <math>T</math></li> <li>the position of words in <math>s_i</math> are in reversed order to those in <math>T</math></li> </ul>
Sentence Reordering	<p><math>s_i</math> is produced by sentence reordering if:</p> <ul style="list-style-type: none"> <li>the words in <math>s_i</math> are found in different sentences in <math>T</math></li> <li>the position of the combined sentences are in reversed order to those in <math>T</math></li> </ul>
Copy-paste	<p><math>s_i</math> is produced by copy-paste if:</p> <ul style="list-style-type: none"> <li>the words in <math>s_i</math> are found in the same sentence in <math>T</math>, say <math>t_j</math></li> <li>the positions of words in <math>s_i</math> are the same as in <math>t_j</math> in <math>T</math></li> <li>the number of word in <math>s_i</math> are equal to <math>t_j</math></li> </ul>

### 3. Summary Sentence Decomposition Algorithms

The main focus of our work is to develop an algorithm to identify students' summarizing strategies using summary sentence decomposition. Summary sentence decomposition is a process to determine whether a summary sentence is generated from the original text and to identify the position of the words in the original text [8]. The task of the summary sentence decomposition algorithm is to:

- determine whether the words in the summary sentence are from the original text,
- locate the locations of the words in the original text using position-based method,
- find the best sequence of locations of words used to represent a phrase in the summary sentence,
- identify the strategies used to produced the summary sentence.

### 4. Conclusions

This paper proposed an algorithm to identify students' summarizing strategies using summary sentence decomposition. Suitable strategies are derived by studying experts' summaries. The strategies are then transformed into a set of heuristic rules. These rules are used to develop an algorithm to identify students' summarizing strategies. We are preparing the evaluation task to determine the efficiency of the algorithm. Teachers find that this is useful to identify their students' skills in summary writing.

### References

- [1] Hidi, S., & Anderson, V. (1986). Producing written Summaries: Task demands, cognitive operations and Implications for instruction. *Review of Educational Research*, 56, 473 - 493.
- [2] Zurina, M. H. (2003). *Summary writing skills by selected form 4 ESL learners utilizing narrative and expository texts*. University Malaya, Kuala Lumpur.
- [3] Zipitria, I., Arruarte, A., Elorriaga, J. A., & Díaz de Ilarraza, A. (2004). *Towards a cognitive model of summary evaluation*. Paper presented at the ITS'2004 on Modeling Human Teaching Tactics and Strategies, Maceió, Brazil.
- [4] Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*, 22, 333 - 362.
- [5] Winograd, P. (1984). Strategic difficulties in summarizing texts. *Reading Research Quarterly*, 19(4), 404 - 425.
- [6] Idris, N, Baba, M. S., & Abdullah, R. (2007). An Analysis on student-written summaries:A step towards developing an automated summarization assessment, *Int. Conf. on Electrical Eng. & Informatics (ICEEI2007)*, 550 - 553.
- [7] Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior*, 22, 1-14.
- [8] Jing, H. (2002). Using Hidden Markov Modeling to decompose human-written summaries. *Computational Linguistic*, 28(4), 527 - 543.