Extractive Text Summarization for Indonesian News Article Using Ant System Algorithm

Abba Suganda Girsang * and Fransisco Junius Amadeus

Computer Science Department, BINUS Graduate Program – Master of Computer Science,
Bina Nusantara University, Jakarta, 11480 Indonesia
*Correspondence: agirsang@binus.edu (A.S.G.)

***Abstract*—The act of simplifying a text from its original source is known as text summarization. Instead of capturing the substance of the original content, an effective summary should be able to convey the information. Recent research on this form of extractive summarization has produced encouraging findings. A graphical model and a modified ant system method will be combined in this literature to provide a solution. The pheromone modification will decide which pertinent phrases will be selected to be a decent summary structure, while the modification process will focus on the point at which the graph construction will be built to represent an article. Additionally, a dataset (Indosum) including news stories that are often utilized in relevant research will be used in accordance to the summary in Indonesian. In addition, the ROUGE approach will be utilized as a tool for evaluation to rate the summary's quality. Finally, this paper concludes with the challenges and future directions of text summarization.**

***Keywords*—ant system algorithm, automatic text summarization, extractive summarization, graph-based summarization**

I. Introduction

The development of the current era forces us to use digital systems to view information sources, one of the most visited are news articles. A digital news text is made up of sentences that contain the essence of the information to be conveyed, and sometimes the information to make it interesting is added with a few sentences to make it look better. This makes the news text very long and takes time to process text by text to understand all the information contained in it. Based on this problem, an automatic summary system has been developed to solve this problem.

The text summarization system, also known as automatic text summarization (ATS), was developed to create a shorter text output that still contains the same information as the source document. ATS is becoming an important topic due to the exponentially increasing amount of textual content on the Internet, while manual digestion of text takes a lot of time, effort and expense and even becomes impractical with a very large amount of textual content [1]. When processing text, ATS has two types of approaches, namely extractive and abstract. The extractive approach selects the key phrases in the input document and then combines them into a summary. The abstraction approach represents the input document in an intermediate view and thus produces a summary with sentences different from the original sentence. This type of extractive summary is often used because it has the advantage of producing a summary without the need to create domain functionality or knowledge, works well as measured by the Rouge score, and provides a coherent summary [2]. Furthermore, the grammatical error rate produced will be lower [3]. Unlike extractive summaries, sentences produced by abstract summaries are new sentences, or commonly called paraphrases, that produce summaries with words that are not in the text. Abstract summaries are very complex and relatively more difficult than extractive summaries because their generation requires extensive natural language processing [4].

TABLE I. Comparison of Extractive Summarization Techniques

Techniques	Advantages	Disadvantage
Statistical based	1. Simple and fast processing. 2. Requires less processor and memory capacity. 3. Unsupervised approaches, no need for training datasets.	No linguistic knowledge processing or semantic relation mapping.
Graph based	1. Can generate query-specific or topic-specific summaries. 2. Unsupervised approaches, no need for training datasets.	Accuracy will rely on the selected affinity function.
Machine-Learning based	1. Simple. 2. Easy to test performance of high number of features.	1. Requires statistical data. 2. Need a huge training corpus for supervised and semi- supe
LSA based	1. Provide Semantic relation. 2. Present important information with least noise.	Difficult to handle polysemy.
Fuzzy-Logic based	1. Knowledge-driven reasoning based, can take better results if integrated with data-driven technique. 2. Fuzzy logic can give compression ratio as low as 20%.	1. Human experts are needed to define the fuzzy rules. 2. Overhead in designing the membership function.

Manuscript received August 23, 2022; revised October 1, 2022; accepted October 12, 2022; published April 4, 2023.

In recent years, there have been many studies that have focused more on extraction approaches. From various studies, it can be concluded that there are five main techniques, namely statistic based, graph based, machine learning based, latent semantic analysis (LSA) based, and fuzzy logic. Research was conducted to draw conclusions on the advantages and disadvantages of each method [5].

Based on the data in the study Table I, several techniques have advantages and disadvantages. Referring to a technique with shortcomings that can still be improved, the graph method has a large segment to develop. The degree of precision based on the function of the method using graphical techniques becomes an important issue. This point is also supported by the many studies that use metaheuristic methods to find algorithmic solutions. This is an important point where combining them can be a new idea worth trying [6, 7]. Several studies have been carried out with metaheuristic algorithms [8–10] and have shown good results in their application.

A useful technique known as Ant Colony has a link with metaheuristic algorithms and graphs where there are parallels between the structure of the graph and the pathways used in it. By making a number of adjustments to the pheromone parameters in the algorithm, this served as the foundation for the introduction of the Ant System algorithm. This enhancement is meant to help the ants take a course that will ultimately result in an effective solution.

II. Related Works

A lot of research has been done on text summaries using graphs because they have weaknesses that can be developed by combining them with other supporting methods. In this case, the graph takes part in the construction of a summary system using a combination of methods. To enable graph-based summaries to be applied, you need to construct an attribute that represents the text and link words or other text entities to meaningful relationships that can be associated with the graph itself.

A study was conducted to prove, using graphics, that their use in ATS has a positive impact on contribution. Using an evaluation method commonly used for text summarization, Recall-Oriented Understudy for Gisting Evaluation (ROUGE) rather than a graph-based idea algorithm (EdgeSumm) obtained the highest ROUGE score in the DUC2001 dataset. For the DUC2002 dataset, the evaluation results show that the proposed framework outperforms the sophisticated ATS system by achieving a 1.2% and 4.7% improvement over the highest scores in the literature for the ROUGE-1 and ROUGE-L metrics, respectively. In addition, the proposed method achieved very competitive results for the ROUGE-2 metric [11]. This proves that the graph method allows improvisation by innovating in the process of building the textual graph, the calculation of the weight of the nodes of the graph, the search graph of the candidate edges, and the algorithms to obtain a candidate summary [11].

Based on the context of graph construction, several weighting methods have been developed, which are used as the basis for nodes and edges in graphs. Research conducted to combine weighting methods and produce an improvisation A major contribution is to produce an improved weighting scheme by combining several important measures that calculate the similarity between two sentences, namely: Jaccard's similarity coefficient, TF*IDF cosine similarity, and the measure identity similarity [12]. This point can be further developed in the process of constructing the graph.

In the process of developing the ATS system with a combination of graphs, other supporting methods are also an important point in making a modification. Further related research tries to do this by presenting an innovative automatic sentence extraction method using graph-based ranking algorithms and the shortest path algorithm [13]. In his research, he came to the conclusion that the shortest path algorithm is better because it produces a smooth summary compared to ranking algorithms. Taking the path from the first sentence to the last sentence does provide a spread to the summary, making it more likely that a significant portion of the original text will be included [13] so that it can represent an idea rather than the document itself. This shows that the Graph Combination algorithm is better than the usual ranking algorithm. In addition, the above statement shows that the use of combinations such as short path algorithms can help. The metaheuristic algorithm itself is widely known for solving short path problems; in this case, the most closely related is a method, namely ant colony. In this case, the similarity between graph and route inspires the idea to develop a graph and ant colony-based extractive summary system, which becomes the proposed method (modified ant system).

III. Proposed Method

Research on the topic of text summarization is still quite interesting. In order to get the best outcomes, a variety of ideas and techniques have been used. This outcome is a problem in and of itself since summary sentences should not only shorten the reading time but also maintain the accuracy of the information they contain with respect to the original text. It is intended that the system summary solution will function as effectively as feasible in this circumstance. The number of studies conducted has become a challenge for how to use the latest methods as a method for solving related problems. The main challenge is to design, adapt, and also create an optimal method that has a high compression ratio without losing the information in it. The deletion of several sentences in the document is a reflection that should be appropriate. In general, the main structure of the proposed method can be described in Fig. 1.

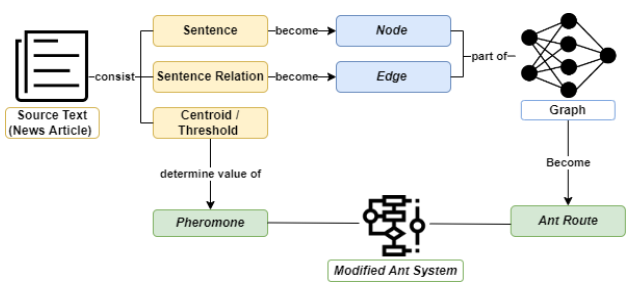

Figure 1. Ant system general concept in text summarization.

The proposed method to find a rough solution focuses on graphs and modification methods of the ant system. By using a graph structure that has similarities with the ant track, it is possible to develop a graph that is related to the summary of news articles. The sentence text can be interpreted as graph using weights, phrases, and sentences in an article to interpret the relationship between them. Moreover**,** by modifying the original ant colony optimization algorithm to a method suitable for summary purposes**,** the main parameter in the form of pheromones is also assigned to news articles by the focal function in the text. This causes algorithm to select relevant sentences whose represents the document based on the threshold or centroid and not the document itself. This centroid itself reflects the essence of the document, as it is obtained by performing the process of averaging all the sentences in the text document. The whole procedure is broken down into steps that can be shown in Fig. 2.

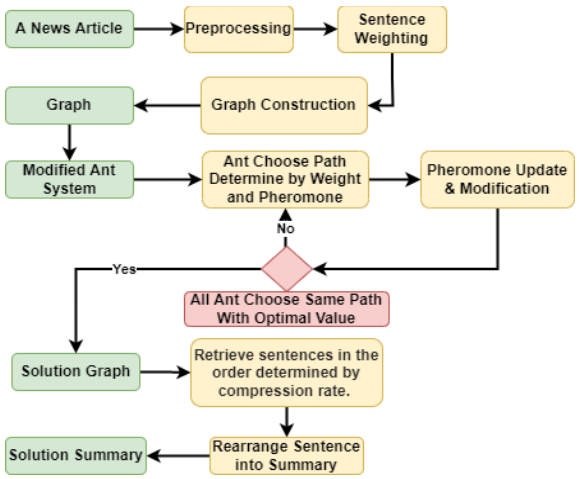

Figure 2. Text summarization process using ant system.

The initial process of summarizing the text begins with the process of changing an article's text into a graph. In this process, it is necessary to preprocess and weight sentences to be able to determine the value of nodes and edges on the graph in the graph construction process. An initial graph that is formed representing an article will be used as a route instead of ants in the modified ant system. In the ant system algorithm process, an iteration process will be carried out to obtain a path that has an optimal value. From a path that is a solution graph, the nth sentence will be taken according to the specified compression rate. In the end, a process of reordering sentences will be carried out because when selecting ants, they tend to choose a random node (sentence) and finally find a solution for the summary results.

A. *Data Preprocessing*

A news article is a dataset that contains several collections of text sentences that have noise. This becomes one of the important points because of the output of a proposed method in processing the dataset. Data that has a lot of noise tend to show poor results where noise will divert the focus of the sentence which will be difficult, especially when weighting will be done. The noise can be in the form of punctuation marks, numbers, special characters, affixes, and so on. Good data in research using text, of course, must go through the preprocessing process. The processes in it include:

- **Case Folding**: This process is useful for removing some noise in the form of capital letters, numbers in words, punctuation marks, and whitespace in text in sentences. All of these are considered unimportant to be processed because they do not have anything relevant to the meaning of a word in summary.
- **Stemming**: Affix is something that is added to a word to make meaning in the sentence so that it can be read and understood better. However, in data processing, there are many affixes in words that are not needed. This is considered to be a differentiator from several words that should be the same. Therefore, in processing text according to studies from previous researchers related to text, it would be better to use basic words. This will make it easier to weight sentences with the next method.
- **Stop Word Removal**: Many irrelevant words in a sentence text. These include words that have no meaning, such as where, when, this, that. Etc. In addition, there are several words in Indonesian that must be excluded so that they do not become noise in word weighting.

B. *Data Weighting and Graph*

An important step in text-related research is weighting. In this study, TF-IDF weighting method will be used to create a vector of sentences. The vector will then be processed using cosine similarity to measure the level of similarity between sentences with other sentences so that a similarity matrix can be obtained. This matrix shows the relations and weights that will relate to each other. Sentence weighting using TF-IDF can be described by Eq. (1).

$$W_{i,j} = tf_{i,j} \times log \frac{N}{df_i} \quad (1)$$

$tf_{i,j} = number\ of occurences\ of\ i\ in\ j$
$df_i = number\ of\ documents\ containing\ i$
$N = total\ number\ of\ documents$

The results of TF-IDF will produce a vector for each sentence where the vector will be used to calculate the weight value of the cosine similarity between the two sentences formulated in Eq. (2).

$$Cos_a = \frac{A \circ B}{|A||B|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \quad (2)$$

$A = Vector\ A, TFIDF\ vector\ of\ the\ sentence\ A$
$B = Vector\ B, TFIDF\ vector\ of\ the\ sentence\ A$
$A \circ B = dot\ product\ between\ vector\ A\ and\ vector\ B$
$|A| = lenght\ of\ A$
$|B| = lenght\ of\ B$
$|A||B| = cross\ product\ between\ |A|\ and\ |B|$

The result of Eq. (2) will be formed in the form of a cosine matrix. This matrix has a relation that describes the nodes and edges of a graph. The graph method is defined by the set of vertices and the set of edges. Vertex represents data entities, and the side represents the relationship between the vertices.

$$G = (N, E) \tag{3}$$

In Eq. (3) describes a graph *G* which has two attributes, namely *N* as a node and *E* as an edge. A collection of relations *E* and *N* will form an undirected graph network where each other will have the relationship described by the edges.

C. Ant System Algorithm

The ant colony algorithm was introduced by Moyson and Manderick *et al.* [14] and extensively developed by Marco Dorigo. This algorithm is a bio-inspired metaheuristic. It has a special group that tries to match the behavioral characteristics of social insects, namely ant colonies. Each actor's behavior is modeled after the behavior of live ants and how they interact with each other to find food sources and efficiently bring them to their colony. Basis of this algorithm was first discovered to solve the Traveling Salesman Problem (TSP), which requires this algorithm to find a solution that has an optimal distance which is influenced by several factors such as the effect of distance and obstacles and how optimal a path is in the process.

The main concept of this algorithm illustrates that the path with the smallest value is not necessarily optimal in its application to the real problem. Because of this, the behavior of ants in search of food has an important factor, namely pheromone and distance. The main illustration of ants finding distance is by utilizing the history of the paths traversed by several ants in several iterations. On the way to reach its destination, an ant will tend to leave a trail, which indicates that the path was traversed by an ant. This element in the ant system algorithm is called a pheromone. The more frequently a path is traveled, the more and larger the amount of pheromone produced. This pheromone itself tends to make ants move at that point because the number of ants that pass will be considered as the optimal path.

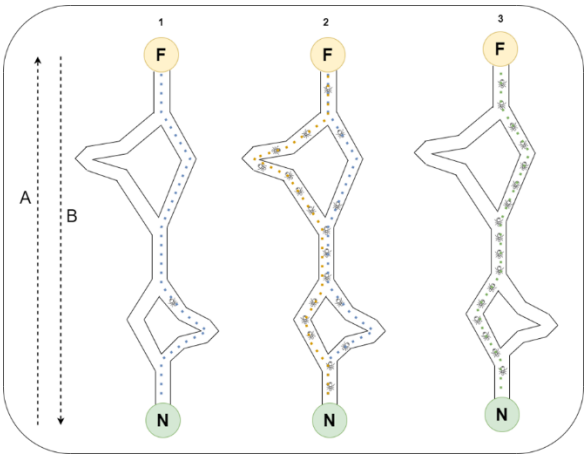

Figure 3. Illustration how ants work in route making decision.

Fig. 3 shows a chronology of how ants find their food: 1) The first ant to find food sources through any route, in this case, is Fig. 3(a). Then return to the nest (N) leaving a trail of pheromone Fig. 3(b); 2) Ants will indiscriminately follow four possibilities, but with stronger pheromones, the ants are more interested in choosing that path as the shortest path; 3) Eventually the ants will take the shortest path, where the other path will increasingly lose the pheromone that has evaporated.

Initially, the ants will be evenly distributed through two paths. After a while, the ants start to choose the shortest path because the pheromones left by the ants will accumulate in the shorter path. An ant k at node i will choose a node j to be in the next layer with probability is defined by Eq. (4):

$$p_{ij}^{A} = \frac{\tau_{ij}^{\alpha} d_{ij}^{\beta}}{\Sigma_k \tau_{ik}^{\alpha} d_{ik}^{\beta}} \tag{4}$$

where denotes the degree of importance of the pheromone and Ni(k) is the choice that ant k (neighborhood) has when it is at node i. The neighborhood of ant k on node i will contain all the vertices that can be reached that are directly connected to node i, except the nodes that have been visited previously. An ant when passing through a segment will leave the number of pheromones contained in a segment in after being passed by an ant k is defined by Eq. (5):

$$\tau_{ij} = (1 - p)\tau_{ij} + \sum_k \frac{1}{L_k} \tag{5}$$

The use of the ant system algorithm in this study is related to the representation of sentences into nodes or points and sentence weights as edges between the distance of one point to another. The relationship between these graphs can be used as the basis for further processing into the path traversed by ants. Based on Eq. (2), a matrix will be formed which can be used to create a graph. The graph used in this study is an undirected graph because we want all the notes to be connected which illustrates that there is a relationship between all sentences. From this relation, the ant system method will find a solution.

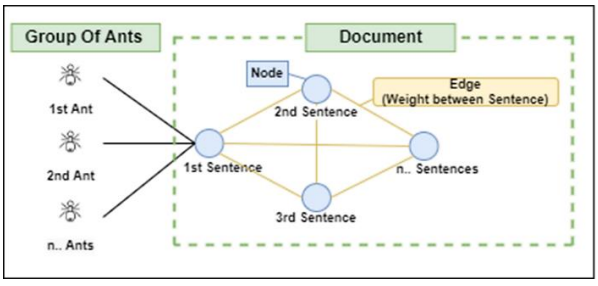

Figure 4. Ant system solution using cosine matrix graph.

Fig. 4 explains how the ant system works. Ants will initially form a colony of n ants. These ants will be run based on certain iterations to find the optimal route rather than the distance in the graph. The cosine matrix is a weight that has a range of 0 to 1 where the larger it is, the more similar it will be. In this case it will be best compared to the behavior of ants where the shortest distance is the

optimal distance. Therefore, there will be a modification to do the inverse where later it can be assumed that the smaller the value or in the edge graph is called the distance, the more similar sentence 1 is to sentence n.

Ants will travel on the graph which is determined by the optimal number of iterations for n ants where each ant will choose a route based on the pheromone and the distance or edge in the graph showed in Fig. 5. The pheromones in this method will experience an enhancement in the form of adjustment to the centroid. The closer a point (sentence) to the centroid, the higher the pheromone will be. This is done to force the ant to choose a point that is relevant later to the core of the document where there will be many choices from one point to another. then Ants may have a different route between ant 1 and ant N, but they tend to have optimal distances and iterations that determine which point is the point at which the algorithm's solution is found. Related research was conducted in determining the number of ants used, stating that the recommended number of ants in the ant system algorithm is between 6 and 20 to achieve optimal results for small to medium ant system applications with an optimal solution or near optimal within a reasonable processing time [15].

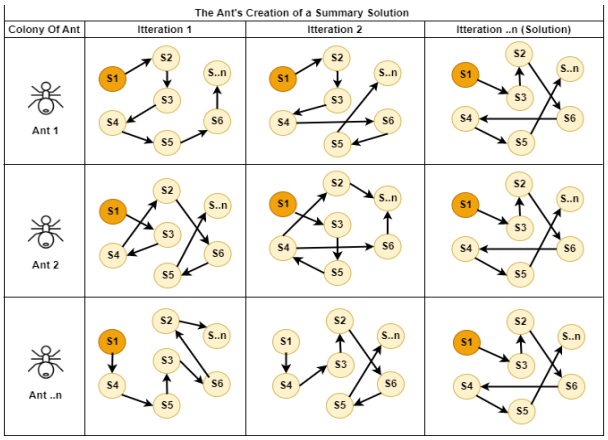

Figure 5. Ant process to find path solution in graph.

Taking the solution will be determined by summing the total distance and the smallest total pheromone that describes an optimal solution of this method. In addition to this, there will be a cut from the route where in terms of summarization there will be a shorter result. This will be determined by the compression rate (CR) parameter. The smaller the CR value, the shorter the resulting summary.

IV. Analysis Results

The method is tested using data from the Indosum dataset. The method developed in this process is unsupervised learning, which focuses on optimizing the search for the most effective results. In this case, the development of process methods in the graphical representation and modification of pheromone values is an important aspect. The experiment was conducted using six categories of news articles, ranging from entertainment, headlines, technology, sports, inspiration, and entertainment. Each article will be processed through stages starting from preprocessing, and weighting to obtain a solution. For the final goal, it will achieve a compression level with a value of 30% of the original sentence according to the results of the reference summary. From every sentence in article, weighting and graphing will be carried out which will produce an indirect graph that has weight. This weight itself has been changed so that the cosine value can be interpreted as a distance so that we can see the distance between sentence 1 and other sentences. The results of the formation and weighting on the graph can be seen in Fig. 6.

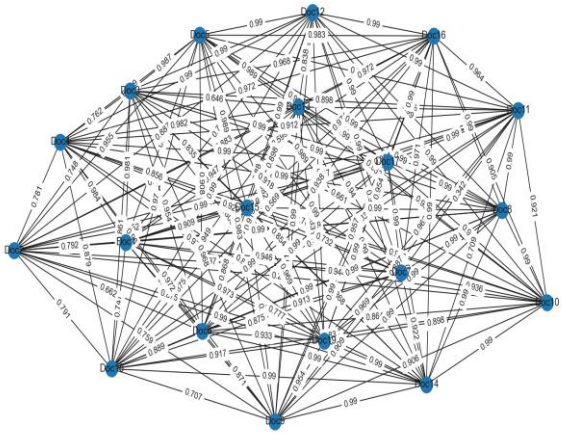

Figure 6. Example of solution graph using cosine matrix graph.

Graph in Fig. 6 is the result of route interpretation where the proposed algorithm makes a solution based on the nature of the ants. The evaluation benchmarks were tested using ROUGE metrics. ROUGE stands for Recall-Oriented Understudy for Gisting Evaluation [16]. It is basically a set of metrics to evaluate automatic text summarization as well as machine translation. It works by comparing an automatically generated summary or translation with a series of reference summaries (usually produced by humans).

- ROUGE-1 is used to evaluate the overlap of 1-gram (every single word) between the system and reference summaries.
- ROUGE-2 is used to evaluate the overlap of bi-grams between the system and reference summaries.
- ROUGE-L is used to evaluate the Longest Common Subsequence between the system and reference summaries

The parameters used by the rogue consist of 3: Precision (*P*), Recall (*R*) and *F-Measure* as shown Eq. (6), (7) and (8).

$$Precision\ (P) = \frac{Relevan_{Sentences} \cap Retrieved_{Sentences}}{Retrieved_{Sentences}} \quad (6)$$

$$Recall\ (R) = \frac{Relevan_{Sentences} \cap Retrieved_{Sentences}}{Relevan_{Sentences}} \quad (7)$$

$$F - Measure = \frac{2 \times Precision * Recall}{Precision + Recall} \quad (8)$$

TABLE II. ANT SYSTEM RESULTS USING ROGUE AND COSINE SIM

	Compression Rate	ROUGE -1	ROUGE-2	ROUGE -L
F-Measure	10%	0.424	0.272	0.336
	20%	0.511	0.378	0.436
	30%	0.602	0.446	0.585
Precision	10%	0.404	0.246	0.330
	20%	0.479	0.367	0.415
	30%	0.606	0.450	0.587
Recall	10%	0.462	0.315	0.378
	20%	0.584	0.409	0.488
	30%	0.652	0.479	0.635

Table II shows the average of the results of each ROUGE value and compression rate tested in this study. There are three values in the results of this measurement: recall shows the relevance of the same number of words compared to the summary reference, precision shows the relevance of the same number of words compared to the summary generated by the system, and F-measure is the level of linkage between the two, which indicates the quality of the system. The results show that the higher the CR, the higher the recall and precision values will be. The three CR values tested (10%, 20%, and 30%) were performed using the dataset, which was mostly calculated to determine the relevance of the value range from that. From these results, it can be seen that the resulting rouge value is good, with the highest value reaching 0.635 at the ROUGE-L value, and shows that the summary has a good value at a compression rate of 30%.

V. CONCLUSION

Based on studies and results, it has been shown that it is still possible to improvise further development of graph-based text summaries. Various methods and combinations of methods can be performed, as in this study. In this case, the research was able to contribute to the algorithmic collaboration process by finding a synthesis that could reflect information from the original document. This is demonstrated by the Rouge value reaching 0.635 on the ROUGE-L with a compression rate of 30%. Based on this research, it can be concluded that the use of extractive summaries is really good and should be applied in real cases to help users read more easily. Further investigation related to the summary may be possible with an abstract method, where it is possible that the words to be summarized are shorter because the system is processing natural language texts, allowing for new words that are shorter than the summary better than the extract method approach.

CONFLICT OF INTEREST

The Author whose names are listed immediately below certify that they have affiliations with Bina Nusantara University in financial interest (such as funding for paper publication) and non-financial interest (such as academic relation) in the subject matter or materials discuss in manuscript.

AUTHOR CONTRIBUTIONS

Abba Suganda Girsang did conceptualization, formal analysis, project administration, validation, writing reviewing, and supervision.

Fransisco Junius Amadeus did data curation, writing-original draft, methodology, analysis, visualization, validation, software, and model development.

REFERENCES

[1] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Syst. Appl.*, vol. 165, 113679, 2021. doi: https://doi.org/10.1016/j.eswa.2020.113679

[2] A. P. Widyassari, P. Adhika, R. Supriadi, and S. Edi, "Review of automatic text summarization techniques & methods," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1029–1046, 2022. doi: https://doi.org/10.1016/j.jksuci.2020.05.006

[3] D. Jani, N. Patel, H. Yadav, S. Suthar, and S. Patel, "A concise review on automatic text summarization," in *Computational Intelligence in Data Mining*, 2022, pp. 523–536.

[4] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: A survey," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 1–66, Jan. 2017. doi: 10.1007/s10462-016-9475-9

[5] A. El-Refaiy, A. R. Abas, and I. Elhenawy, "Review of recent techniques for extractive text summarization," *J. Theor. Appl. Inf. Technol.*, vol. 96, pp. 7739–7759, Dec. 2018.

[6] P. Verma and A. Verma, "A review on text summarization techniques," *J. Sci. Res.*, vol. 64, pp. 251–257, 2020. doi: 10.37398/JSR.2020.640148

[7] P. Verma and H. Om, "A novel approach for text summarization using optimal combination of sentence scoring methods," *Sādhanā*, vol. 44, no. 5, p. 110, 2019. doi: 10.1007/s12046-019-1082-4

[8] H. Asgari, B. Masoumi, and O. S. Sheijani, "Automatic text summarization based on multi-agent particle swarm optimization," in *Proc. 2014 Iranian Conference on Intelligent Systems (ICIS)*, 2014, pp. 1–5. doi: 10.1109/IranianCIS.2014.6802592

[9] D. A. Prabowo, M. Fhadli, M. A. Najib, H. A. Fauzi, and I. Cholissodin, "TF-IDF-enhanced genetic algorithm untuk extractive automatic text summarization," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 208–215, 2016.

[10] N. Chatterjee, A. Mittal, and S. Goyal, "Single document extractive text summarization using genetic algorithms," in *Proc. 2012 Third International Conference on Emerging Applications of Information Technology*, pp. 19–23, 2012. doi: 10.1109/EAIT.2012.6407852

[11] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "EdgeSumm: Graph-based framework for automatic text summarization," *Inf. Process. Manag.*, vol. 57, no. 6, 102264, 2020. doi: https://doi.org/10.1016/j.ipm.2020.102264

[12] A. Alzuhair and M. Al-Dhelaan, "An approach for combining multiple weighting schemes and ranking methods in graph-based multi-document summarization," *IEEE Access*, vol. 7, pp. 120375–120386, 2019. doi: 10.1109/ACCESS.2019.2936832

[13] K. S. Thakkar, R. V. Dharaskar, and M. B. Chandak, "Graph-based algorithms for text summarization," in *Proc. the 3rd Int. Conf. Emerg. Trends Eng. Technol. ICETET 2010*, 2010, pp. 516–519. doi: 10.1109/ICETET.2010.104

[14] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: Optimization by a colony of cooperating agents," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 26, no. 1, pp. 29–41, 1996.

[15] M. M. Alobaedy, A. A. Khalaf, and I. D. Muraina, "Analysis of the number of ants in ant colony system algorithm," in *Proc. 2017 5th International Conference on Information and Communication Technology (ICoIC7)*, 2017, pp. 1–5. doi: 10.1109/ICoICT.2017.8074653

[16] C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," in *Proc. Work. Text Summ. Branches out (WAS 2004)*, 2004, pp. 25–26.

Abba Suganda Girsang obtained the Ph.D. degree in the Institute of Computer and Communication Engineering, Department of Electrical Engineering and National Cheng Kung University, Tainan, Taiwan, in 2014. He graduated with the bachelor from the Department of Electrical Engineering, Gadjah Mada University (UGM), Yogyakarta Indonesia, in 2000. He then continued his master's degree in the Department of Computer Science in the same university in 2006–2008. He joined the Faculty of Department of Informatics Engineering in Janabadra University as a lecturer in 2003–2015. He also taught some subjects at some universities in 2006–2008. His research interests include swarm intelligence, business intelligence, machine learning and media social text mining.

Fransisco Junius Amadeus received the bachelor's degree in computer science study program of Bina Nusantara University Jakarta, Indonesia. Currently he is studying for master's degree at Bina Nusantara University, Jakarta, Indonesia. He has freelance experience developing a web-based system solution. He is currently a senior member of staff and the project leader for system development projects in the core insurance business division at Sinar Mas Insurance Company in Jakarta, Indonesia.