

Extractive Text Summarization of Student Essay Assignment Using Sentence Weight Features and Fuzzy C-Means

I Made Suwija Putra^{a,1,*}, Yonatan Adiwinata^{a,2}, Desy Purnami Singgih Putri^{b,3}, Ni Putu
Sutramiani^{a,4}

^a Department of Information Technology, Faculty of Engineering, Udayana University, Bali

^b Graduate School of Department of Electrical Engineering and Computer Science, Kanazawa University, Kanazawa,
Japan

¹ putrasuwija@unud.ac.id*; ² natanzt1@gmail.com; ³ desysinggihputri@stu.kanazawa-u.ac.jp; ³ sutramiani@unud.ac.id
* corresponding author

ARTICLE INFO

Article history:

Received 03-08-2020

Revised 23-10-2020

Accepted 06-01-2021

Keywords:

Text Summarization

Essay Assignments

Weight Sentences

Fuzzy C-Means

ABSTRACT

One of the main tasks of a lecturer is to give students an academic assessment in the learning process. The assessment process begins with reading or checking the answers of student assignments that contain a combination of very long sentences such as essay or report assignments. This certainly takes a lot of time to get the primary information contained therein. It is necessary to summarize the answers so that the lecturer does not need to read the whole document but is still able to take the essence of the response to the task. This study proposes the application of summarizing text documents of student essay assignments automatically using the Fuzzy C-Means method with the sentence weighting feature. The sentence weighting feature is used by selecting the sentence with the highest weight in one cluster, helping the system to get the primary information from a document quickly. The results of this study indicate that the system succeeds in summarizing text with an average evaluation of the values of precision, recall, accuracy, and F-measure of 0.52, 0.54, 0.70, and 0.52, respectively.

Copyright © 2021 International Journal of Artificial Intelligence Research.

All rights reserved.

Introduction

The rapid development of information technology makes a massive increase in the number of digital texts. According to MGI research [1] that in 2010, people in the world stored more than 6 exabytes of digital text data on devices such as personal computers (PCs), notebooks, and mobile phones. Digital text is text in digital format in the form of a representation of the binary alphabet [2]. It can be found in article documents, news, books, scientific papers, or student assignments collected online. These documents often contain long text, so it requires time to filter the information in it. One example, it takes a lot of time for lecturers to give a proper assessment of student essay assignment documents. This because they have to read all the documents to get the core information. Text summarization becomes one of the quick and appropriate solutions to find the core sentences of the essay document so that it can speed up the assessment process.

Summarizing the text is to condense it into a shorter version while maintaining the primary information and its overall meaning. It is challenging to summarize large text documents manually [3]. Therefore, this study proposes an automatic digital text summarization system to summarize student essay assignment documents and display essential sentences. This study uses the technique of extracting materials extensively with the Fuzzy C-Means method, which is to give sentence weight in each sentence the student answers.

The Fuzzy C-Means method, developed by Dunn in 1973 and improved by Bezdek in 1981, is often used in pattern recognition [4]. Fuzzy C-Means can make each object a member of a cluster of several choices smoothly, able to group extensive data, more robust against outlier data, and effective and efficient. Besides that, it is simple and easy to implement [5]–[8].

Previous studies on summarizing documents used as a basis for this research include Research [9], which implements Fuzzy C-Means to summarize English-language journal documents. The application created can summarize with input pdf format text. Sentence weighting uses the TF IDF method.

Research [10] developed a text summarization system model consisting of four stages. The preprocessing step changes the unstructured text into the structured one.

Other research [11] proposed the topic of sentence-based Bayes models to summarize documents using terms and associations. An efficient variational Bayesian algorithm is derived for estimating model parameters. Experimental results on benchmark data sets indicate the effectiveness of the proposed model for multi-document summarizing tasks.

Other studies [12] propose making automatic document summaries using word features and the K-Means method. Automatic document summaries can be used to get text summaries quickly, making it easier for users to get the primary information from a document. This study summarizes documents using word features and the K-Means method.

Previous studies generally use documents such as articles and news which have no multi-topic content and use different languages. As for materials such as essay assignment documents or student scientific papers consisting of several questions and answers, at this time, the summarizing has not been done much. Based on these reasons, research on the automatic summarizing of texts for student essay assignment documents using the Fuzzy C-Means method and weighting sentences (TF ISF) be raised. This research used the data from the assignment of the Udayana University Department of Information Technology students. All of the data are in Indonesian.

Method

Research on Extractive Text Summarization of Student Essay Assignment Using Sentence Weight Features and Fuzzy C-Means has the following initial stages: (1) Conduct a literature study on the concept of summarization and the Fuzzy C-Means method; (2) The collection of test data sourced from digital texts of student essay assignments collected through e-learning systems; (3) Preparation of test data before processing by the system includes data acquisition, data cleaning and conversion of file formats to .txt.

After these stages are carried out, the next step is to design the system flow design and implement the plan. Fig. 1 shows the flow of the system starting from the text of the document to be summarized; of course, it has been through the stage of preparing test data. In the next step, the test data enters the summarization system that begins with indexing, which is the process of finding all the features in the text, both word features, punctuation features, numeric features, and others in the document.

Furthermore, this study has two main processes, namely the summarization process and the results of the summarization evaluation process. The Summarization with extractive method has three general procedure steps, as below[13][14]

Step 1: The first step is making a document representation. Preprocessing techniques performed in this level, including tokenization, stop word removal, stemming, sentence separation, frequency calculation, and others.

Step 2: The sentence assessment is being done in this step. Three approaches are followed:

- Weighting of words to determine which are the important words by using the TF-IDF method
- Weighting of sentences such as verifying sentence features (their position in the document, similarity to titles, sentences containing important words) by using the TS-ISF method.
- Graph scoring to analyze the relationships between sentences.

Step 3: In this step, the sentences are sorted, then take the highest score as the final summary in a single document

The Summarization Process on this study as presented in Fig. 2 has several stages, such as Indexing, Clustering Process with Fuzzy C-Means, and Sentence Weighting in Clusters. The system evaluation process is comparing results from the summary of the system with the manual review by the expert.

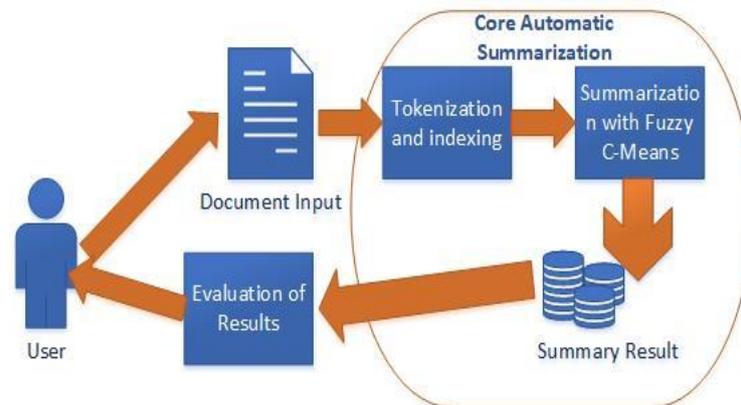


Fig. 1 Overview system of Extractive Text Summarization of Student Essay Assignment Using Sentence Weight Features and Fuzzy C-Means

A. Indexation

The indexing process starts from inputting the .txt file as text to be summarized into the system. This process begins by separating each sentence. Separation of sentences is done by using several indicators such as periods (.), Exclamation points (!), And question marks (?). Examples of difficulties encountered are the use of period punctuation (.), it is not only used when ending a sentence but also used to abbreviate names and others. In this process, case folding is carried out to change all the letters into a uniform form. In this study, all words uniformed in lower cases. Furthermore, words are taken in each sentence (tokenization) and discard the words that are the words that appear most often but do not have significant meaning (stop word). The list of stop words taken from the Indonesian stop words database. Next is the stemming process, which is the process of removing affixes to get essential words in Indonesian. This system uses a library from Sastrawi [15].

B. Clustering Sentences with Fuzzy-C-Means Method

The sentence clustering process is needed to organize extensive data collections by partitioning several data sets automatically, so objects that have similarities will be grouped into a group that is different from other groups [5][16].

The results of the preprocessing stage are used to form clusters that contain sentences whose features are close together. Basically, the Fuzzy C-Means (FCM) algorithm has a lot in common with K-Means. The output of FCM is not a fuzzy inference system, but a row of cluster centers and several membership degrees for each data point [17]. The FCM grouping procedure is as follows:

1. Define the following parameters.

- a. Number of clusters = c (adjusted according to the level of summarization)
- b. Powers of number = w
- c. Maximum iteration = $MaxIter$
- d. Error threshold value = ξ
- e. Initial objective function P_0 = 0
- f. Initial iteration = $t = 1$

- Initialize the membership data as matrix form, X , with the size of $n \times m$ (n = number of data samples, m = data attributes), where X_{ij} = index of data sample ($i = 1, 2, \dots, n$) and attributes ($j = 1, 2, \dots, m$). The membership value derived from the frequency of words in each sentence.
- Generating random values from 0 to 1 for each member of the group, μ_{ik} , $i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$; as an elements of initial partition matrix u . μ_{ik} . Calculate the total number of each column (attribute) with the following formula:

$$Q_j = \sum_{k=1}^c \mu_{ik} \quad (1)$$

Q_j is degrees of membership per column = 1 with $j=1, 2, \dots, m$, then calculate with the following formula:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j} \quad (2)$$

- Calculate the centroids of k : V_{kj} , with $k = 1, 2, \dots, c$; and $j = 1, 2, \dots, m$ and value of μ_{ik} from formula 2.

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik}^w) * X_{ij})}{\sum_{i=1}^n \mu_{ik}^w} \quad (3)$$

- Calculate the value of the objective function in the t iteration, P_t . The objective function is used to get the right cluster center. It means the *cluster* of data in last iteration was achieved.

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] \mu_{ik}^w \right) \quad (4)$$

$$i = 1, 2, \dots, n; k = 1, 2, \dots, m$$

- Calculate the difference of membership level for each data in partition matrix:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-\frac{1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-\frac{1}{w-1}}} \quad (5)$$

with: $i=1, 2, \dots, n$; and $k=1, 2, \dots, c$.

- Iteration end if the difference of objective function value less than the Error threshold value ($|P_t - P_{t-1}| < \xi$) or the number of iterations has passed the maximum iteration limit ($t > \text{MaxIter}$). If the above condition not met, $t = t + 1$ and repeat to step 4.

C. Sentence Weighting with TS-ISF

Term frequency weighting - inverse sentence frequency (TF-ISF) done after getting all the iteration process completed. Before searching for the TF-ISF value, the word weight value must be found using the TF-IDF method [17]. This method is a numerical value to represent how importance the words is in the whole document. TF shows word frequency in documents. IDF is a measure to reduce the weight of words that often appear in the corpus and increase the weight of words that rarely occur.

At this stage, the TF-ISF value of each sentence will be added up and used as the value of a sentence, which will be used at the sentence selection stage in each cluster. The ISF and TF-ISF equations can be seen in the following equations number 1 and 2 [18]:

$$ISF_t = \log\left(\frac{N}{SF_t}\right) \tag{1}$$

$$TFISF_{t,s} = TF_{t,s} \times ISF_t \tag{2}$$

With $TF_{t,s}$ is the frequency of occurrence of the word t in the sentence s , N is the number of sentences in the document, and ISF_t is the number of sentences containing the word t . The $TFISF_{t,s}$ value, s , will be high if the word t appears several times in a sentence and rarely appears in another sentence, and low if the word t appears almost in the entire sentence [19].

The final stage of the summarization process is to take the sentence that has the highest sentence weight for each cluster. The number of sentences taken as a result of summarization depends on the selected concise level value, which starts from 10% -50%. For example, if there are 20 sentences in the text tested with a summarization rate of 40%, the system will create 8 clusters (the user selected), then one sentence which is the closest to its center will be taken from each cluster.

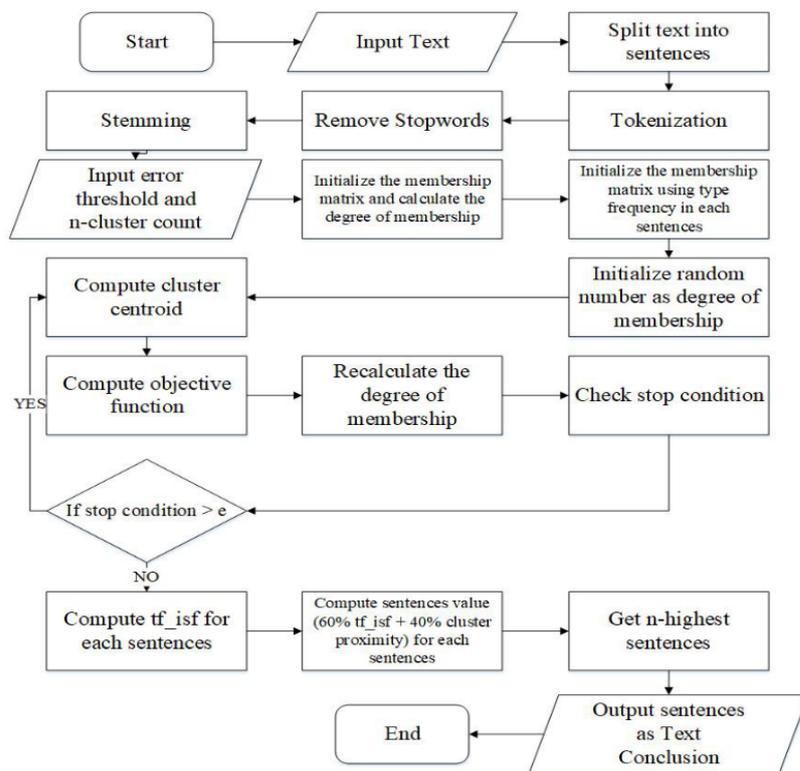


Fig. 2 Algorithm of the text summarization system using the Fuzzy C-Means method and weighting of sentences

D. System Evaluation Techniques

The process carried out in the system evaluation stage is to compare the results of summarization by the system with the manual version. There are several evaluation techniques in measuring the quality of the sentence grouping model, including information matrix, misclassification index, purity, f-measure [20]. This study uses the F-Measure Technique in order to get the value of accuracy, precision, and recall, and f-measure of the summary results issued by the system, this is used as an indicator of the results of this research. The flow of the system evaluation process in this study is shown in Fig. 3.

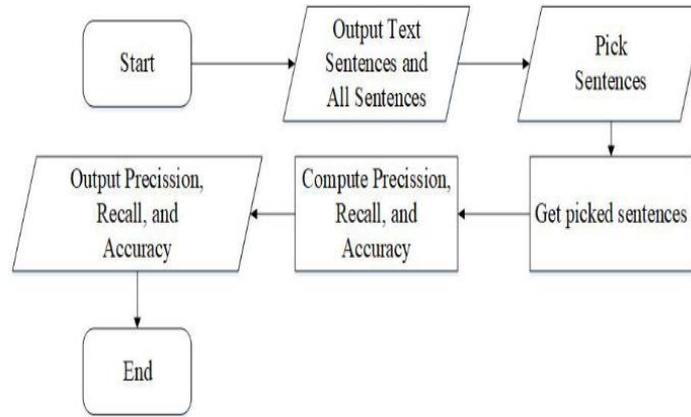


Fig. 3 System evaluation process flowchart

The f-measure measurement is based on the precision and recall values obtained. The concepts of recall and precision can be seen in Table 1 below.

Table 1 The concept of counting recall and precision

	Relevant	Non-Relevant
Retrieved	tp	fp
Not Retrieved	fn	tn

The recall is the proportion of sentences rediscovered as a summary, and precision is the proportion of the number of sentences found and considered relevant [21]. From Table 1, the formula for calculating recall, precision, accuracy, and f-measure values described as follows.

$$Recall (R) = \frac{tp}{tp+fn} \tag{3}$$

$$Precision (P) = \frac{tp}{tp+fp} \tag{4}$$

$$F - Measure = \frac{2RP}{R+P} \tag{5}$$

$$Accuracy = \frac{tp+tn}{tp+fp+fn+tn} \tag{6}$$

True positive (tp) is a sentence that is in the expert summary and appears in the system summary. False positives (fp) sentences that are not in the expert summary but appear in the system summary. False-negative (fn) is a sentence that is in the expert summary but does not appear in the system summary. And true negative (tn) is a sentence that is not in the experts and system summary.

Result and Discussion

The summarization system for student essay assignment documents automatically uses the Fuzzy C-Means method built using the Python programming language with WX packet as the system interface. The overall process interface of this system can be seen in Fig. 4.

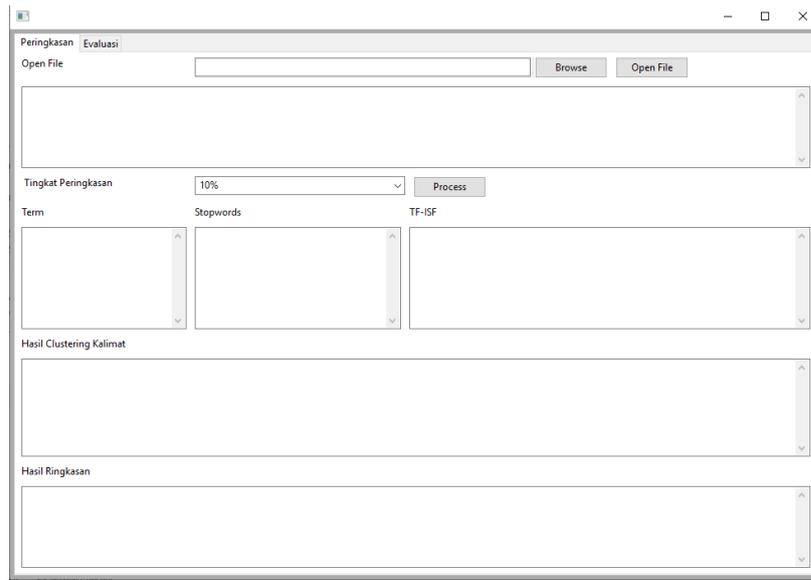


Fig. 4 General System interface of the summarization system for student essay assignment documents automatically uses the Fuzzy C-Means

A. Test Data Collection

The document test data used in this study was obtained from student assignments collected in the e-learning system (ELSE U) at Udayana University in 2019. Document collection is done by downloading randomly without regard to any indicators. The student essay document was achieved from 10 student in the e-learning system with the same assignment topic.

The entire document was selected by looking at the length of the students answers to make it more reliable when tested. The selection process gives ten documents, which are then converted manually by copy-pasting each sentence in the PDF document in accordance with the rules that have been made into a .txt format file.

This collection process takes a long time because each sentence must be checked and adjusted to the rules that have been made i.e., a collection of sentences must be complete, more than three sentences in 1 paragraph, and not a sentence that explains the tables, pictures or formula equations. The results of the conversion resulted in an average of 113 words omitted in each document. The comparison of the number of words after and before the conversion can be seen in Fig. 5 below.

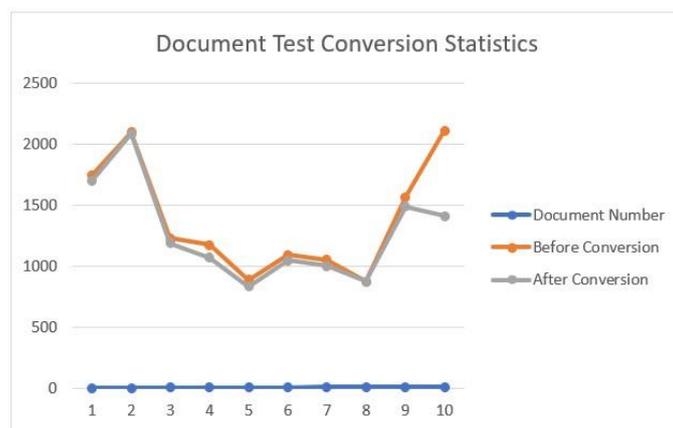


Fig. 5. Change in the number of words when converting test documents to .txt format. The x-axis is the document number, the y-axis is the number of words

As shown in Fig. 5, the average of reduction from 10 document is 6,87%. The deleted words such as the non-standard word and word that describe objects, pictures, tables and equations or

formulas because these words are not reliable. So, the remaining 93.13% of the document content is included in the process of converting documents or test data. Test data also used as a manual summarization material conducted by experts to be used as a comparison with the results of a summary of the system to carry out the evaluation process. In this study, it assumed that the results of the manual review are good. Fig. 6 shown the result of testing data that was processed by format conversion and word reduction as mention before.

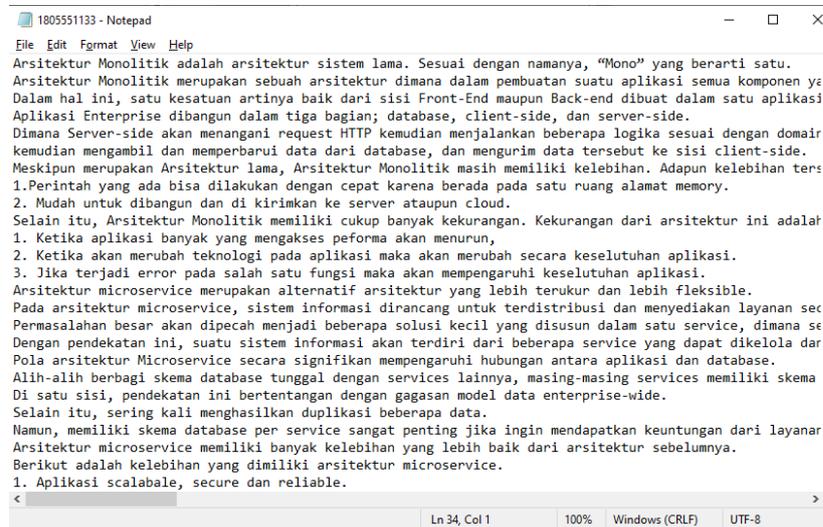


Fig. 6 Example of testing data that converted to .txt format

From 10 documents that have been converted, paragraphs that have a minimum 4 sentences are selected, so that each n cluster produced by the Fuzzy C-Means clustering process can be filled with correct members, that will have selected 1 sentence for every 1 cluster[22]. Fuzzy C-Means clustering is also useful for researcher to create variations for ideal extractive summaries [23]. From the process of taking this paragraph, it produces 105 text which will be the test data to be processed in testing the automatic text summarization system using this Fuzzy C-Means method.

B. Implementation of Text Summarization

System implementation is the stage after carrying out the system design process to complete the system to be ready for use and test functionality is done to make sure the algorithm of Fuzzy C-Means with the TF-ISF sentence weighting can produce clusters containing representing sentences to become a summary based on the summary level chosen. This study uses the complete parameters of the Fuzzy C-Means method for this experiment, namely the error threshold (the smallest error expected) = $1,6 \times 10^{-5}$, maximum iteration = 100, and $(w)= 2$.

Fig. 7 displays a system flow diagram that describe program flow from input document until the output process. The first stage of this summarizing process is to input the file that will be summarized by selecting the test data file that has been prepared. In the second stage, choose the summary rate with a range from 10% to 50%. This level of summarization affects the results of summarization, where the smaller the level of summarization chosen, the number of sentences produced will also be less.

The third stage, when the user presses the process button, will get the results in the form of a list of the frequency of words contained in the document and stop word obtained from the test data. The next step is to calculate the results of the weighting sentences to determine the ranking in each cluster. Sometimes some clusters do not have members because the number of sentences tested is not enough to form the cluster that determined at the beginning. If this happens, then the sentence that will have the second-highest closeness of each cluster has members up to the number of sentences selected according to the value of the rank level. The final stage of this system is to bring up the summarized results from the tested inputs.

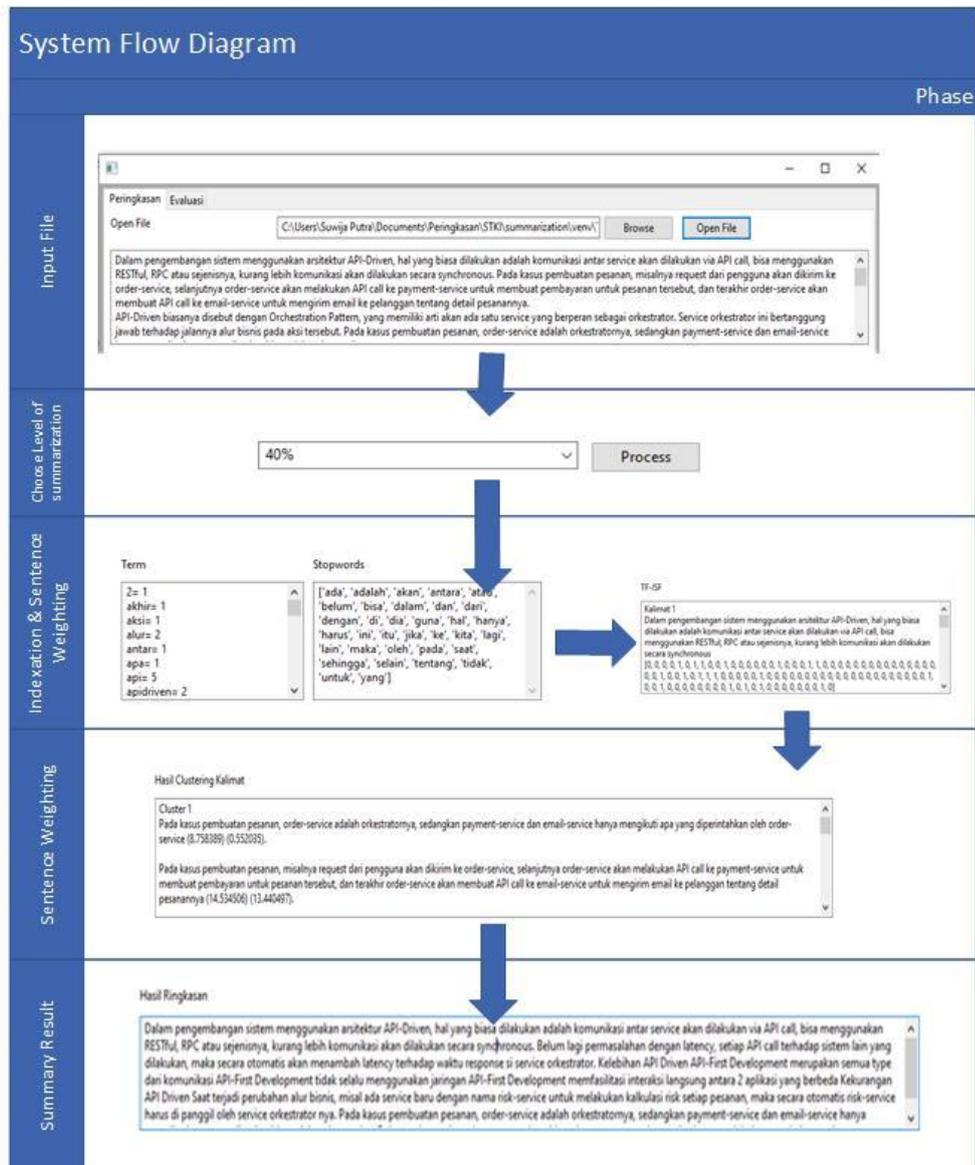


Fig. 7. System flow diagram of the summarization system for student essay assignment documents automatically uses the Fuzzy C-Means

C. System Evaluation Results

The system evaluation in this study aims to measure the success rate of summary results issued by the system compared to manual summaries by experts. Experts do the manual review by summarizing the contents of documents based on their knowledge to choose which sentences are considered as good as summaries.

Testing is done with a concise level of 10%, 20%, 30%, 40% and 50% for each data. This value is said to be sufficient because a summary is said to be good enough if the summary results are not more than 50% of the entire contents of the automatic text document on the system interface that has been provided [3]. The number of test data is 10.

Table 2. The results of the average value of precision, recall, accuracy and F-Measure based on a concise level of 10-50%

Compression Rate	Indicator			
	Precision	Recall	Accuracy	F-Measure
10%	0,31	0,33	0,78	0,32
20%	0,40	0,43	0,71	0,41
30%	0,59	0,61	0,71	0,59

40%	0,59	0,62	0,64	0,61
50%	0,71	0,68	0,67	0,69
Average	0,52	0,54	0,70	0,52

The graph in Fig. 8 shows that a compression rate of 10% has a high accuracy value, but the other value is very low. This means that there is refraction, that is, sentences that should be of high weight cannot be brought up by the system. The best results in this study were obtained at a compression rate of 50%, although the accuracy value decreased the value of precision and recall increased so that the bias that occurs as a compression rate of 10% becomes smaller.

Table 2 shows the average results of the system evaluation tests obtained. This indicates that the Fuzzy C-Means method succeeds in summarizing the text in documents by using the sentence weight feature (TF-ISF) as an indicator of the closeness of the sentence to the center of the cluster. Evaluation values obtained are precision values of 0.52, recall of 0.54, the accuracy of 0.70, and f-measure of 0.52. If the average values of precision, recall, accuracy, and f-measure are compared, they look like in Fig. 8.

The evaluation carried out in this study is very dependent on the manual summary. Manual summary based on the wishes of each expert to choose which sentence is considered suitable as a summarization, this results in different results if the experts are different.

The higher level of summarization makes the number of sentences that come out as a result of the system summary also increases. Meanwhile, the number of summary sentences issued by the system and the good manual summary also greatly influences the value of tp, fn, fp, and tn. The highest accuracy value in this study obtained from the lowest level of summarization; this is because the value of tn and tp tends to decrease in the summary results that produce a large number of sentences (high summarization rate value). Vice versa, if the level of summarization is low, then the number of summary sentences produced will be less so that the accuracy value tends to be higher.

From the evaluation that have been done, the Fuzzy C-Means with sentence weighting has a fairly good accuracy in processing cluster formation of data, and also proves that the Fuzzy C-Means method is dependent on the type of data. If the data character is linear with many dimension, Fuzzy C-Means will get faster performance and better accuracy [24] [25].

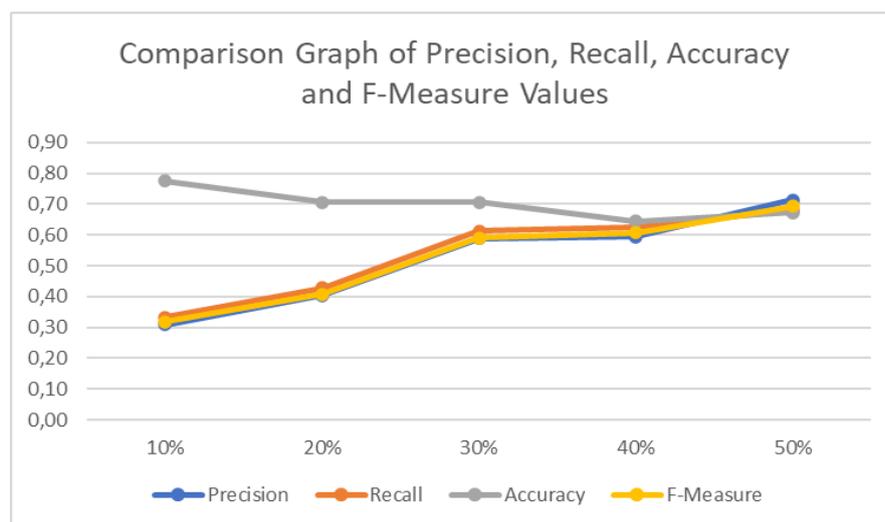


Fig. 8. Comparison of average precision, recall, accuracy, and F-Measure values

Conclusion

Based on the results of the trials conducted, it can be concluded that the clustering text using Fuzzy C-Means makes the summary results change every time the experiment so that the evaluation result level of the summary always changes or is not static because taking the initialization cluster

center point is done randomly. In addition, the results of system evaluations also depend on the quality of the manual summaries produced by experts. The results of the system evaluation of 10 test data get an average value of the overall level of compression (compression rate) for the precision of 0.52, recall of 0.54, the accuracy of 0.70, and f-measure of 0.52. These results indicate that the fuzzy c-means performance with TF-ISF sentence weighting that used to produce extractive text summaries cannot be said to be optimal, in further studies it can be developed by applying an optimization method that works with the Fuzzy C-Means method.

Acknowledgment

Thank you to Udayana University for providing support in the form of grants for this research, as well as those who helped to complete this research.

References

- [1] B. A. Manyika J, Chui M, Brown B, Roxburgh C, "Big data: the next frontier for innovation, competition, and productivity," [MGI] McKinsey Glob. Inst., 2011.
- [2] N. O. Finnemann and N. O. Finnemann, "E-text," in *Oxford Research Encyclopedia of Literature*, Oxford University Press, 2018.
- [3] V. Gupta and G. S. Lehal, "A Survey of Text Summarization Extractive techniques," in *Journal of Emerging Technologies in Web Intelligence*, Aug. 2010, vol. 2, no. 3, pp. 258–268, doi: 10.4304/jetwi.2.3.258-268.
- [4] R. Bharathi, S. C. Shirwaikar, and V. Kharat, "A distributed, scalable parallelization of fuzzy c-means algorithm," 2016, doi: 10.1109/IBSS.2016.7940196.
- [5] V. K. Singh, N. Tiwari, and S. Garg, "Document clustering using K-means, heuristic K-means and fuzzy C-means," 2011, doi: 10.1109/CICN.2011.62.
- [6] B. Adnyana, "Implementasi Algoritma Fuzzy C Means Dan Statistical Region Merging Pada Segmentasi Citra," *Konf. Nas. Sist. Inform. 2015*, pp. 9–10, 2015.
- [7] K. S. Gilda and S. S. Dixit, "Clustering : Basics , Approaches , Practical View and Applications," *Int. J. Comput. Eng. Appl.*, vol. X, no. Vii, pp. 19–29, 2012.
- [8] W. Wang, C. Wang, X. Cui, and A. Wang, "Fuzzy C-means text clustering with supervised feature selection," in *Proceedings - 5th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2008*, 2008, vol. 1, pp. 57–61, doi: 10.1109/FSKD.2008.161.
- [9] M. Irfan, Jumadi, W. B. Zulfikar, and Erik, "Implementation of Fuzzy C-Means algorithm and TF-IDF on English journal summary," in *Proceedings of the 2nd International Conference on Informatics and Computing, ICIC 2017*, Feb. 2018, vol. 2018-January, pp. 1–5, doi: 10.1109/IAC.2017.8280646.
- [10] R. Al-Hashemi, "Text Summarization Extraction System (TSES) Using Extracted Keywords," *Int. Arab J. e-Technology*, vol. 1, no. 4, pp. 164–168, 2010, [Online]. Available: [http://www.iajet.org/iajet/iajet_files/vol.1/no.4/Text Summarization Extraction System TSES Using Extracted Keywords_doc.pdf](http://www.iajet.org/iajet/iajet_files/vol.1/no.4/Text%20Summarization%20Extraction%20System%20TSES%20Using%20Extracted%20Keywords_doc.pdf).
- [11] D. Wang, S. Zhu, T. Li, and Y. Gong, "Multi-Document Summarization using Sentence-based Topic Models," 2009. doi: 10.3115/1667583.1667675.
- [12] M. R. Muztahid, "Peringkasan Dokumen Bahasa Indonesia Menggunakan Metode K-Means," 2015.
- [13] A. Nenkova and K. McKeown, *Automatic summarization*, vol. 5, no. 2–3. 2011.
- [14] S. Akter, A. S. Asa, M. P. Uddin, M. D. Hossain, S. K. Roy, and M. I. Afjal, "An extractive text summarization technique for Bengali document(s) using K-means clustering algorithm," 2017, doi: 10.1109/ICIVPR.2017.7890883.
- [15] H. A. Robbani, "Sastrawi," *MIT*, 2016. .
- [16] M. Ozer, "Fuzzy c-means clustering and Internet portals: A case study," *Eur. J. Oper. Res.*, vol. 164, no. 3 SPEC. ISS., pp. 696–714, 2005, doi: 10.1016/j.ejor.2003.11.015.
- [17] M. Afsharizadeh, H. Ebrahimpour-Komleh, and A. Bagheri, "Query-oriented text summarization

- using sentence extraction technique,” 2018, doi: 10.1109/ICWR.2018.8387248.
- [18] Z. Ghahramani, “Probabilistic machine learning and artificial intelligence,” *Nature*. 2015, doi: 10.1038/nature14541.
- [19] C. D. Manning, P. Raghavan, and H. Schutze, *Introduction to Information Retrieval*. 2008.
- [20] I. Yoo and X. Hu, “A comprehensive comparison study of document clustering for a biomedical digital library MEDLINE,” 2006, doi: 10.1145/1141753.1141802.
- [21] K. Vimal Kumar and D. Yadav, “An improvised extractive approach to hindi text summarization,” in *Advances in Intelligent Systems and Computing*, 2015, vol. 339, pp. 291–300, doi: 10.1007/978-81-322-2250-7_28.
- [22] V. N. Phu, N. D. Dat, V. T. Ngoc Tran, V. T. Ngoc Chau, and T. A. Nguyen, “Fuzzy C-means for english sentiment classification in a distributed system,” *Appl. Intell.*, vol. 46, no. 3, 2017, doi: 10.1007/s10489-016-0858-z.
- [23] D. A. Riandayani, I. K. G. Darma Putra, and P. W. Buana, “Comparing Fuzzy Logic and Fuzzy C-Means (FCM) on summarizing indonesian language document,” *J. Theor. Appl. Inf. Technol.*, vol. 59, no. 3, 2014.
- [24] Z. Rustam, A. Kamalia, R. Hidayat, F. Subroto, and S. A. Suryansyah, “Comparison of Fuzzy C-Means, Fuzzy Kernel C-Means, and Fuzzy Kernel Robust C-Means to classify thalassemia data,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 9, no. 4, 2019, doi: 10.18517/ijaseit.9.4.9580.
- [25] Z. Rustam and A. S. Talita, “Fuzzy kernel K-medoids algorithm for multiclass multidimensional data classification,” *J. Theor. Appl. Inf. Technol.*, vol. 80, no. 1, 2015.