

INTRODUCTION

Text summarization is a technique of processing a text document with a computer program to produce a summary that retains the essential essence of a text document [1]. In this globalization era, the use of text summarization is increasingly widespread. By using text summarization it is easier for someone to get information by just reading a text summary [6][7].

Text summarization in general has two text summarizing approaches which are classified into extractive and abstractive. Extractive summarization is a system's process of copying the main features of the original document and merging into a shorter version. Abstractive summarization is generated from extracting the original document and then generated by adding new sentences that are different from the previous original document[6][9]. The extractive approach has the characteristics of simple text summarization. This simplicity makes the extractive approach more efficient than the abstractive approach[1][2][4].

Latent Semantic Analysis (LSA) is a method of sequence-based machine learning. The choice of LSA method for extractive summarization is because LSA can be used for term similarity machines based on hidden topics and can be used for clustering [5]. LSA is also well known for its success in summarizing using an extractive approach[2][5]. However, this method has a weakness that its measurement is only based on the relationship between all sentences, which makes the larger the sentence the required supporting features that have greater intelligence[2][5][7].

LSA has several ways to solve text summarization, one of which is the TF-IDF method[5][7]. TF-IDF assigns each word weighting value in the summarized document[2][10]. TF-IDF is often used in factor weighting or weighting in information retrieval[5][10].

According to Makbule Gulcin Ozsoy, Ilyas Cicekli and Ferda Nur Alpaslan in the Gong & Liu summary algorithm, the first concept is selected, and then the sentence most related to this concept is selected as part of the resulting summary. Then the second concept is selected, and the same steps are executed. Repetition selects the concept and sentences most related to that concept continues until a predetermined number of sentences is extracted as part of the summary[11].

In the research of Makbule Gulcin Ozsoy, Ilyas Cicekli and Ferda Nur Alpaslan, it was also explained that the SteinbergerJezek method works by choosing sentences that relate to all important concepts and at the same time choosing more than one sentence from an important topic[11].

Cross is an enhanced result of the SteinbergerJezek method. The cross method works by processing the summary results not only based on the similarity of words and sentences in a document, but also the length of the sentence determines the success of increasing the precision value of the LSA F measurement in previous studies, that's what has been done[11].

Each summary should be scored to see how good the summary results are. Initially, summary evaluation is done manually with reference to the human evaluation of the summary results. However, manual evaluation has some drawbacks, such as assessments that are not objective because everyone has different standards in assessing a summary. In addition, manual evaluation is associated with a lot of time and effort. To address this weakness, an investigation is conducted for the automatic implementation of summary evaluation. And one of the results of this study is ROUGE, which is from ChinYew Lin[12].

Elements of ROUGE elements include:

Rouge N

ROUGE N is a unigram, bigram, trigram and higher order n-gram overlap between the candidate summary and the specified reference summary. Ngram in RougeN is the number of word lengths in an abstract candidate that will be compared to the reference to the reference to the reference[12].

Rouge L

Rouge L works by measuring a group of words using the longest match with LCS. The advantage of using LCS is that it does not require sequential matches, but sequential matches that reflect the order of words at the sentence level. Because it automatically contains general programs in the longest order, you do not need a predetermined program duration[12].

Rouge S.

With skip bigram metrics, these metrics are a collection of multiple pairs of words in a sentence. You can search for sequential words in reference text that appear in the output of the model but are separated by one or more words. ROUGE S allows us to add some leeway to our n gram match[12].

Based on the research described above, this study presents a summary of Indonesian news documents using the Gongliu, Steinberger Jazek and Cross method, which aims to create a system that is able to provide important information in summary form with quick results. Then the results of the text summarization will be compared with the three methods with the comparison parameters obtained from the Rouge score and the time during extractive automatic text summarization processing takes place.