

CHAPTER I

INTRODUCTION

1.1 Background

Along with the development of current technology, various types of information can be easily found on the internet, one of them is a document in the form of text data. For short text data with a short period of time, the contents can be known, but for long text data, the entire text in the document must be reviewed in order to understand its contents, this requires a long time and effort. Another problem that arises is when you have to understand a topic in depth, it is necessary to review not only one text document but many documents at the same time, reading the entire document will require a long time and also a great effort to be able to understand its contents. The solution is to read a summary of the document, by reading the summary only requires less time than when you have to read the entire contents of the document, and the effort needed to understand the contents is also less.

To overcome this problem, a method of reading documents can be done at a glance (skimming). But this sometimes makes some information on a document difficult to understand because the information may require other information that was previously unreadable due to skimming. Besides skimming, another method that can be done is to read a summary of a document (summary). The summary is a representation of a document that contains the essence or main focus of the

document. A summary can increase the effectiveness of the reader in finding and finding the desired document [1]. However, a document generally does not have a summary because making a summary of a document is quite time-consuming and costly [2]. In addition, summarizing documents is not easy because a summary must be able to represent the entire contents of the document.

Based on the background of the problems described, a document summarizing system is needed that is able to automatically summarize (automatic summarization) from a document to facilitate users in receiving information.

1.2 Statement of the Problem

Based on the background described above, there are several problems that can be formulated as follows:

1. How much accuracy does the system produce using the Maximum Marginal Relevance algorithm in producing document summaries in Indonesian?

1.3 Scope and Limitation

The system limits that will be made in this study are as follows:

1. The text used for testing is the Indonesian text.
2. The assumptions of Indonesian language documents to be summarized have language patterns that refer to (in Indonesian statement) called Ejaan Yang Disempurnakan (EYD) or standard.

3. Does not handle word writing errors.
4. The abbreviation is considered one word.
5. Summarizing this document only applies to a single document.
6. Documents that are input are .doc and .docx files with the condition that the contents are writing not images.
7. The algorithm used in this program is MMR (Maximum Marginal Relevance).
8. The dataset is available (original text and summary).

1.4 Objectives of the Research

The objectives to be achieved by the author are:

1. Produce a system that can summarize Indonesian text using the Maximum Marginal Relevance algorithm.
2. Know how effective the Maximum Marginal Relevance algorithm is in summarizing Indonesian documents.

1.5 Benefits of the Research

By utilizing this system, it is expected that it will be easier for users to get broad outline information about the desired document. The system can help the reader, absorb all the information contained in the document, because it produces a new text but still contains the most important parts of the original text through calculations to produce predetermined accuracy.

1.6 Research Methods

This study uses the Descriptive Method [3] which aims to get a clear picture of things that are needed. In this text summarization study, the data studied is as it is. So the text summarization results depend on the data itself.

1.6.1 Literature Review

Literature studies are carried out by studying theories through books, articles, journals and other materials related to the stage method used to summarize the text and the Maximal Marginal Relevance algorithm.

1.6.2 Data Collection

In this study, the method used is the study of literature. Study of literature obtained from written sources, both printed and electronic, such as books, e-books, journals, papers, and sources related to the problems studied.

1.6.3 System Development

This stage is the stage of making the program in accordance with the design that has been made. First is the creation of a database that will be used to store word dictionaries and stopword lists that have been downloaded before. The two programs that are able to upload files will be summarized and produce a summary.

1.6.4 System Testing and Analysis

Testing of the program is done by evaluating the summary results generated by the system. Then an analysis of the summary results of the system will be carried out and measures the effectiveness of the programs that have been built.

1.7 Outline

The systematics of writing the Akhir Task report is grouped into 5 chapters, namely:

Chapter I Introduction. This chapter contains an overview of the research that will be carried out which includes background, problem formulation, problem boundaries, research objectives, research methods, and systematics of research writing.

Chapter II, Theoretical basis. This chapter contains a literature review and theory used in this study. The literature review outlines various theories obtained from various sources related to this research.

Chapter III, System Design. This chapter contains the system design that will be built in this study. The system design will be made in the form of specifications of the system, the design of the system diagram, the design of the system interface in the form of input and output.

Chapter IV, System Implementation. This chapter is about the implementation of the system that has been designed as planned in chapter III.

Chapter V, *Conclusions and Suggestions*. This chapter contains conclusions from the results of the research that has been done and suggestions that might be made for the development of further research.

