ABSTRACT

Nowadays, the development of technology continues to increase, thanks to this progress a technology called the internet was born. With the internet, everyone can easily connect with the outside world. All information can be accessed on the internet easily, everyone in Indonesia is connected to the internet. One of the most popular search engines used in the world is Google, with billions of users each visit. One of Google's popular channels for researchers to publish research results is Google Scholar. Google Scholar was created to make it easier for users to find or use educational reference materials and scientific research in a publication format. This reference can be in the form of scientific journals which are the results of research, research or studies that can be justified scientifically.

In this final project, a research which will use the K-Means Clustering method is conducted. This method is one of the grouping or grouping algorithms that means learning without supervision, which means that the input of this algorithm accepts data without class labels. The function of this algorithm is to group data into clusters. The data will be grouped by the K-Means method, previously it will be crawled first from Google Scholar. Then the data that can be filled in the file will be processed by extracting text to generate cloudwords from keywords in the document.

Data from Google Scholar managed to crawl with the contents of the pdf document, then the document will be input into the R shiny application so as to produce a word representation of all documents and keywords from the abstract that were most successfully displayed in cloudword, the results of the presentation of accuracy using the K-Means method Clustering in this final project will also display Cluster Plot graph visualization.

Keywords: K-Means Clustering, Crawling Data, Data Mining, Google Scholar, Internet History, Google History