ABSTRACT

Nowadays, the development of technology continues to increase, thanks to this progress is born a technology on behalf of Internet. With the internet everyone can easily connect with the outside world. All information can be accessed easily, every day millions of people in Indonesia are connected to the Internet. One of the most popular search engines used in the world is Google, with billions of users every day. One of Google's popular channels for researchers to publicize the results of his research is Google Scholar. Google Scholar is created to make it easier for users to search for or browse educational materials or references to scientific research in the form of publications. This reference can be a scientific journal which is the result of a scientifically accountable research, research or study.

In this final project, a study will be conducted that will use the Analysis of Variance (ANOVA) method. a statistical analysis that examines the differences in the average data of more than two groups. The function of this algorithm is as an analytical tool to test research hypotheses which assess whether there are average differences between groups. Data that will be grouped with the Analysis of Variance (ANOVA) method, will be crawled beforehand from Google Scholar.

Data from Google Schoolar was successfully crossed with pdf document content, then the scientific publication document was inputted into the R shiny application for further analysis using the Analysis of Variance (ANOVA) method to produce keyword representations of all documents tested. With the creation of this system, it is expected to be a test tool to find out how representative keywords are in scientific documents with the entire contents of the document.

Keyword: *Analysis of Variance, data crawling, data mining, google scholar*