ABSTRACT

The quality and quantity of data has a large influence on the results of research and makes data collection an important role in the research process. Images data that will be obtained from the results of scraping on google images is the original link, as well as all images, and also the results of relations that will be obtained from the results of the keywords entered.

Selenium webdriver supports taking data contained in google images that are dynamic. Retrieval of data from google images is expected to use the selenium library, is expected to get all relationships obtained on google images, then combined with itertools, then calculate the storage needed by using linear regression and also the results of the original links in accordance with the keywords to be searched.

Engine that is made has several features, among others: get all image lines without being limited, combine keywords with categories, and predict the amount of storage size needed. The results of the comparison of the size of the storage size required for repeat 1 is 95,84% repeat 2 is 95,55%, repeat 3 is 84,69% and finally repeat 4 is 55,06%. Features that have been made to run with their respective functions.

***Keywords**: Scraping, Google Images, Selenium Webdriver, Relation, Keyword, itertools, linear regression.*