

ABSTRAK

Kualitas dan kuantitas data memiliki pengaruh besar dalam hasil penelitian dan membuat pengumpulan data sebagai peran penting dalam proses penelitian. Data gambar yang akan di peroleh dari hasil *scraping* di *google images* adalah tautan aslinya, juga seluruh gambar yang terdapat di *google images*, dan ada pula hasil relasi yang akan di peroleh dari hasil keyword yang di masukan.

Selenium webdriver mendukung pengambilan data yang terdapat di *google images* yang bersifat dinamis. Pengambilan data dari *google images* diperkirakan akan menggunakan *library selenium*, diperkirakan akan mendapatkan seluruh relasi kategori yang di peroleh pada *google images*, lalu dikombinasikan dengan *itertools*, kemudian menghitung penyimpanan yang dibutuhkan dengan menggunakan regresi linear dan juga hasil tautan asli sesuai dengan kata kunci yang akan dicari.

Engine yang dibuat memiliki beberapa fitur, antara lain : mendapatkan seluruh tautan gambar tanpa dibatasi, mengkombinasikan keyword dengan kategori, dan memprediksi besaran ukuran *storage* yang dibutuhkan. Hasil perbandingan besaran ukuran *storage* yang dibutuhkan pada *repeat* 1 sebesar 95,84%, *repeat* 2 sebesar 95,55%, *repeat* 3 sebesar 84,69% dan terakhir *repeat* 4 sebesar 55,06%. Fitur yang telah dibuat berjalan dengan fungsinya masing-masing.

Kata kunci: *Scraping, Google Images, Selenium, Relasi, Keyword, Itertools, regresi linear.*