

Abstrak

Web crawler merupakan salah satu komponen *search engine* yang berperan penting dalam melakukan penelusuran web. Berbeda dengan *crawler* biasa, *focused crawler* berusaha menjaring web yang sesuai dengan topik dengan melibatkan *classifier* di dalamnya. Dari beberapa *crawling strategy*, algoritma *fish search* memiliki keunggulan dalam hal akurasi. Jika proses pencarian web dianalogikan ke dalam sebuah *graph*, algoritma *fish search* bekerja dengan terlebih dahulu mengecek relevansi *node* awal dengan topik, kemudian dilanjutkan menelusuri *node* anak. Pengujian dilakukan untuk mengetahui performansi berupa akurasi, presisi, dan waktu proses. Skenario uji yang dilakukan diantaranya dengan membandingkan URL yang terjaring antara *focused crawler* dengan *crawler* biasa, mengubah komposisi *dataset training*, penggunaan *pre-processing*, penambahan jumlah *dataset training*, serta mengubah nilai *depth*. Dan diperoleh hasil bahwa *focused crawler* lebih baik dalam menjaring URL yang sesuai dengan topic, diperoleh komposisi dataset training terbaik untuk masing-masing pengujian, *pre-processing* meningkatkan akurasi, dan penambahan *depth* berbanding lurus dengan kenaikan akurasi.

Kata Kunci : *Focused Crawler, Fish Search, dataset training.*