Abstrak

Perkembangan world-wide web yang semakin pesat diikuti oleh kebutuhan informasi yang semakin meningkat, menjadi tantangan yang belum pernah terjadi sebelumnya untuk *general-purpose crawler* dan *search engines*. *Search engine* seperti Google, Yahoo!, Altavista dan sebagainya telah diperkenalkan dan digunakan untuk mempermudah pencarian informasi di Internet. *Web Crawler (crawler)* adalah sebuah *program/script* otomatis yang memproses halaman web untuk sebuah mesin pencari, yang banyak digunakan saat ini. Namun mengingat banyaknya halaman web yang ada, maka seringkali *search engine* dengan *crawler* biasa tidak dapat memberikan hasil yang maksimal. Untuk itu dikembangkanlah *focused crawler*. *Focused crawler* akan men-*download* halaman web yang sesuai topik dan berhati-hati memutuskan URL mana yang akan di-*scan* dan dalam urutan apa dilanjutkan berdasarkan informasi halaman download sebelumnya. Untuk tugas akhir ini, yang akan diproses adalah web musik.

Focused crawler membutuhkan classifier untuk membedakan halaman web yang relevan dan tidak. Yang pada tugas akhir ini digunakan Naïve Bayes Classifier. Halaman web yang relevan akan diekstrak outgoing-linknya dan disimpan kedalam *frontier*. Link dapat dicrawl dengan menggunakan algoritma penelusuran. Pemilihan algoritma penelusuran yang tepat akan berpengaruh pada efisiensi web crawler. Pada tugas akhir ini yang akan digunakan adalah *Learning Anchor Algorithm*.

Berdasarkan implementasi, dihasilkan akurasi terbaik 100% pada link pengujian http://gigsplay.com/ dengan dataset 100 musik dan 100 nonmusik. Sedangkan akurasi terendah 86.7% saat dataset 300 musik dan 200 nonmusik pada link pengujian http://musik.kapanlagi.com/.

Kata Kunci : focused crawler, web olahraga, naïve bayes, *Learning Anchor Algorithm*.