

Abstrak

Focused crawler adalah crawler yang khusus mendownload halaman web yang sesuai dengan topik yang ditentukan oleh pengguna. Tugas utama dari focused crawler adalah mengumpulkan sebanyak-banyaknya halaman web yang relevan sesuai dengan topik yang telah ditentukan. Tidak semua halaman web didownload dalam sebuah web site, tetapi hanya halaman web yang berkaitan dengan topik saja yang akan disimpan, sehingga dapat menghemat penggunaan resource dari server.

Tugas Akhir ini mengimplementasikan sebuah focused crawler dengan menggunakan metode cosine similarity, link score, dan traverse irrelevant page. Metode cosine similarity digunakan untuk menentukan apakah sebuah halaman web dinyatakan sesuai dengan topik yang dicari. Metode link score digunakan untuk memandu crawler ke arah mana kira-kira akan mendapatkan halaman web yang relevan dengan topik. Metode traverse irrelevant page merupakan teknik penelusuran halaman web yang tidak relevan untuk mendapatkan halaman web yang relevan di dalam-nya.

Hasil pengujian menunjukkan bahwa focused crawler akan mendapatkan nilai precision rate yang optimal dengan menggunakan metode traverse irrelevant page dengan kedalaman level 0. Focused crawler ini juga dapat diimplementasikan dengan menggunakan seed url yang keterkaitannya dekat dengan topik maupun seed url yang hanya sedikit keterkaitannya dengan topik. Performansi dari focused crawler dilihat dari parameter precision rate dan waktu komputasi juga akan optimal jika menggunakan seed url yang sedikit keterkaitannya dengan topik.

Kata Kunci: *focused crawler, cosine similarity, link score, traverse irrelevant page, weight table.*