

1. Pendahuluan

1.1 Latar belakang

Pencarian dengan menggunakan mesin pencari telah menjadi hal yang umum dan sangat dibutuhkan oleh masyarakat pada saat ini. Pencarian dilakukan dan diharapkan menghasilkan hasil yang jauh lebih tepat dan akurat. Pencarian tidak lagi hanya didasarkan oleh kemiripan *query* tetapi juga mulai didasari atas adanya rekomendasi dari penulis dokumen halaman *web* yang lain.

Proses pencarian berdasarkan prinsip *text-based analysis* didasari atas isi teks yang ada di dalam suatu dokumen [8]. Proses pencarian berdasarkan prinsip ini memiliki kekurangan yaitu untuk mencari dokumen yang relevan maka *user* dituntut untuk memberikan inputan *query* yang secocok mungkin dengan *term* yang ada di dalam dokumen. Oleh karena itu, munculnya mesin pencari dengan menggunakan prinsip *link analysis*. Prinsip pencarian berdasarkan *link analysis* ialah proses pencarian dan pengurutan dokumen halaman *web* berdasarkan informasi yang ada di dalam *link* tersebut atau disebut juga adanya rekomendasi dari penulis dokumen halaman *web* yang lain [8].

Algoritma SALSA adalah algoritma yang akan dibahas dalam Tugas Akhir ini. Algoritma SALSA (*Stochastic Approach for Link-Structure Analysis*) merupakan salah satu jenis algoritma *link analysis*. Algoritma ini dikembangkan oleh Lempen dan Moran dimana dalam menentukan halaman web yang *relevan* menggunakan prinsip *hub* atau *authority* [1,3,8,10]. *Authority* ialah nilai yang dimiliki oleh suatu dokumen karena dituju oleh dokumen halaman *web* lain [1,3,8,10]. *Hub* ialah nilai yang dimiliki suatu dokumen karena mengacu ke dokumen lain [1,3,8,10]. Penentuan pengurutan tertinggi didasari atas nilai *authority* dimana semakin besar nilai *authority*-nya maka semakin tinggi pengurutannya.

1.2 Perumusan masalah

Adapun masalah yang akan diselesaikan pada Tugas Akhir ini, yaitu:

1. Bagaimana pengaruh jumlah *link* yang dimiliki oleh suatu dokumen terhadap pengurutan *kerelevanannya*?
2. Bagaimana pengaruh jumlah dokumen yang ter-*crawler* untuk menentukan pengurutan dokumen – dokumen halaman *web*?
3. Bagaimana pengaruh *single document* dan *multi document* dalam proses pengurutan dengan menggunakan Algoritma SALSA?
4. Bagaimana pengaruh jumlah *root set* dalam pengurutan dokumen halaman *web*?

1.3 Batasan Masalah

Adapun batasan-batasan yang diberikan dalam penyelesaian masalah Tugas Akhir ini adalah sebagai berikut :

Batasan-batasan masalah pada Tugas Akhir ini adalah :

1. Penelitian pada Tugas Akhir ini berfokus pada Algoritma SALSA (*Stochastic Approach for Link-Structure Analysis*).
2. Proses *crawling* dilakukan secara *On-line*.

3. Proses perhitungan Algoritma SALSA terhadap dokumen ter-*crawling* dilakukan secara *offline*.
4. Sistem tidak menangani pencarian *root set* berdasarkan *text-based analysis*.
5. Sistem tidak menangani pembobotan *term* pada proses *searching*.

1.4 Tujuan

Tujuan Tugas Akhir ini adalah sebagai berikut.

1. Menganalisis pengaruh jumlah *link* yang dimiliki oleh suatu dokumen halaman *web* dalam pengurutan dokumen halaman *web* tersebut.
2. Menganalisis pengaruh banyaknya jumlah dokumen yang *tercrawling* dalam penentuan pengurutan dokumen halaman *web*.
3. Menganalisis pengaruh *single* dan *multi document* terhadap pengurutan dokumen halaman *web*.
4. Menganalisis pengaruh banyaknya jumlah *root set* terhadap hasil pengurutan dokumen halaman *web*.

1.5 Metodologi penyelesaian masalah

Metode yang digunakan untuk menyelesaikan permasalahan-permasalahan Tugas Akhir ini terdiri dari langkah-langkah sebagai berikut.

1. Pengumpulan data dan studi literatur
Mempelajari dan memahami salah satu algoritma *link analysis* yaitu algoritma SALSA melalui literatur berupa buku, makalah, atau jurnal dari berbagai media terutama Internet.
2. Pembangunan Aplikasi
Pembangunan aplikasi yang meliputi:
 - a. Perancangan Sistem
Menyiapkan dokumen yang akan digunakan untuk dilakukan proses pengurutan berdasarkan algoritma SALSA. Proses selanjutnya dilakukan perancangan sistem yang nantinya dapat menghasilkan suatu sistem yang dapat melakukan pengurutan dokumen halaman *web* berdasarkan jumlah *link* yang dimiliki oleh dokumen halaman *web* tersebut.
 - b. Implementasi
Di tahap ini dilakukan implementasi algoritma SALSA pada dokumen yang diperoleh, yang meliputi:
 - Memproses sejumlah *root set* yang diberikan untuk dilakukan proses *crawling* terhadap sejumlah *root set* tersebut.
 - Setiap *root set* yang *dicrawling* dilakukan proses penyimpanan *link* yang dimiliki oleh *root set* tersebut.
 - Proses perhitungan jumlah *link* yang *tercrawling* berdasarkan algoritma SALSA.
 - c. Pengujian
Sistem aplikasi yang sudah dibangun kemudian diuji untuk mengetahui apakah sistem sudah berjalan seperti yang diharapkan.

- 1) Untuk menganalisis efektifitas aplikasi, pengujian dilakukan dengan menjalankan aplikasi *search engine* dan kemudian dilakukan percobaan dengan menggunakan *single document* dan *multi document*.
 - 2) Untuk menganalisis pengaruh jumlah dokumen dan jumlah *root set* maka dilakukan pengujian dengan memperbanyak jumlah atau nilai keduanya apakah dapat menghasilkan hasil pengurutan yang berbeda.
- d. Analisis Hasil
- Pada tahap ini didapat *output* dari hasil perhitungan algoritma SALSA kemudian dihitung nilai *IAP* dan *precision* berdasarkan prinsip *backlink metric* dan dianalisis apakah algoritma SALSA tepat untuk proses perangkaan.
- e. Penyusunan Laporan
- Pada tahap ini, akan dilakukan penyusunan laporan akhir sekaligus dokumentasi dengan mengikuti kaidah penulisan yang benar dan sesuai dengan ketentuan yang ditetapkan oleh institusi.