

1. Pendahuluan

1.1 Latar belakang

Dengan berkembangnya teknologi yang ada, maka berkembang pula kebutuhan akan informasi. Oleh sebab itu, banyak dokumen informasi yang tersedia. Banyaknya dokumen yang tersedia menimbulkan permasalahan yakni sulitnya menemukan dokumen informasi yang dibutuhkan secara tepat dan cepat. Dengan banyaknya jumlah dokumen informasi, dibutuhkan waktu yang lama untuk menemukan informasi-informasi yang relevan dengan kebutuhan *user* dari sekian banyak kumpulan dokumen. Oleh karena itu, pada Tugas Akhir ini penulis mencoba mencari dokumen yang relevan berdasarkan prinsip *link-based analysis*. Pada prinsip *link-based analysis*, dokumen yang memiliki topik yang sama dapat merekomendasikan dokumen lain begitu juga sebaliknya. Dalam *link-based analysis* dikenal istilah *authoritative* yang berarti suatu dokumen tidak hanya relevan tetapi juga populer [12]. Dengan *link based analysis*, *user* dapat memperoleh dokumen yang tidak hanya relevan dengan kebutuhannya namun juga populer atau dapat menemukan dokumen yang benar – benar dapat menjadi sumber yang informatif.

Pada Tugas Akhir ini akan dilakukan penelitian mengenai *Authoritative Dokumen Halaman Web* dengan menggunakan algoritma HITS (*Hyperlink-Induced Topic Search*). Algoritma HITS merupakan algoritma yang diciptakan untuk membantu menemukan dokumen yang *authoritative* dengan memanfaatkan struktur *link* dalam pencarian dokumen [1,7]. Seorang pencipta halaman *web* akan memberikan *link* lain pada halaman *web*-nya yang dapat memberikan informasi yang sama seperti yang ia miliki. Dalam Algoritma HITS akan dilakukan penghitungan nilai *authority* dan *hub* dalam proses pencariannya. *Authority* ialah dokumen yang memiliki topik serupa dengan topik yang *user* berikan. *Hub* berarti dokumen halaman *web* yang memiliki *link* ke *authority pages* [8,11]. Perhitungan nilai *authority* dan *hub* ini akan dilakukan secara *iterative* hingga diperoleh nilai yang *konvergen*. Setelah diperoleh nilai yang *konvergen*, dokumen yang memiliki nilai *authority* yang tinggi merupakan dokumen yang *authoritative* dengan kebutuhan *user* [11].

Oleh sebab itu, dalam Tugas Akhir ini akan diteliti bagaimana pencarian *authoritative* dokumen dengan menggunakan algoritma HITS. Dengan Algoritma HITS, diharapkan dapat memberikan hasil yang berkualitas baik dan mengembalikan dokumen sesuai dengan kebutuhan *user*.

1.2 Perumusan masalah

Masalah yang akan diselesaikan dalam Tugas Akhir ini adalah sebagai berikut.

1. Bagaimana efektivitas algoritma HITS dalam menemukan dokumen yang *authoritative* dilihat dari *precision* dan IAP?
2. Bagaimana pengaruh jumlah dokumen ter-*crawling* dan *root set* terhadap dokumen yang dihasilkan dari hasil pengujian yang dilakukan?

1.3 Batasan Masalah

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

1. Penelitian pada tugas akhir ini hanya fokus pada satu algoritma saja yaitu algoritma HITS (*Hyperlink-Induced Topic Search*).
2. Dalam melakukan *crawling* (penelusuran link) bersifat *online*.
3. Tidak menangani *preprocessing* yaitu proses pencarian dengan *text based* untuk *root set*.
4. Dokumen yang di-*download* merupakan dokumen *web*.

1.4 Tujuan

Tujuan Tugas Akhir ini adalah sebagai berikut.

1. Menganalisis efektivitas algoritma HITS dalam menemukan dokumen yang *authoritative* dengan mengamati *precision* dan IAP.
2. Menganalisis pengaruh jumlah *root set* dan batasan jumlah dokumen yang di-*crawling* terhadap dokumen hasil pengujian.

1.5 Metodologi penyelesaian masalah

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

1. Studi Literatur
Mempelajari dan memahami algoritma HITS melalui literatur berupa buku, makalah, atau jurnal dari berbagai media terutama Internet.
2. Perancangan dan Pemodelan Sistem
Pada tahap ini akan dirancang sebuah *information retrieval system* yang didalamnya dapat melakukan perhitungan nilai *authority* dalam menemukan dokumen yang relevan dengan topik yang diberikan.
3. Implementasi Sistem
Pada tahap ini dilakukan pembangunan terhadap *system* yang telah dirancang. Pada tahap ini akan dibangun *system* yang dapat menangani proses *crawling*, proses perhitungan nilai *authority* dan *hub*, dan proses *searching*.
4. Pengujian dan analisis hasil.
Pengujian dilakukan untuk memperhatikan performansi Algoritma HITS dalam *information retrieval*. Pada proses ini akan dilakukan perhitungan nilai *precision* dan IAP.
5. 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.