1. PENDAHULUAN

1.1 Latar Belakang Masalah

Saat ini, *web* telah menjadi sarana pencarian data maupun informasi penting yang dibutuhkan oleh masyarakat. Dengan banyaknya data yang tersebar dalam seluruh jaringan internet itu sendiri, perlu dipilah bagian *web* yang mengandung informasi penting atau tautan menuju informasi penting lainnya. Dalam proses pemilahan atau yang sering disebut pemeringkatan *web*, berperan sangat penting dalam sebuah sistem pencarian informasi. Proses ini akan menghasilkan rekomendasi *web* yang dianggap penting, biasanya dapat dihitung dari banyaknya *web* yang mengacu ke *web* tersebut. Diharapkan dengan banyaknya *web* yang mengacu *web* tersebut, menandakan *web* tersebut memiliki informasi yang penting.

Pemeringkatan ini dilakukan dengan prinsip *link analysis*, yaitu proses pemeringkatan dokumen berdasarkan informasi yang terkandung di dalam *link* dari sebuah halaman *web*. *Link analysis* adalah salah satu dari banyak faktor yang dipertimbangkan oleh mesin pencari *web* dalam komputasi skor komposit untuk halaman *web* pada setiap *query*-nya[6]. *Link analysis* telah menunjukkan potensinya dalam peningkatan kinerja pencarian dokumen *web*.

Terdapat dua algoritma pemeringkatan yang populer digunakan, yaitu PageRank dan HITS [2][4][6]. PageRank dan HITS merupakan sebuah algoritma yang berfungsi untuk menentukan situs *web* yang lebih penting atau populer. Dalam penerapannya, algoritma HITS ditekankan pada hubungan yang saling menguatkan pada halaman *web*, sedangkan pada PageRank ditekankan pada normalisasi berat dengan menggunakan model *random walk*[2][6]. Sebagian besar algoritma *link analysis* yang ada, memberlakukan halaman *web* sebagai satu *node* tunggal. Namun dalam berbagai kasus, halaman *web* berupa semantik sehingga tidak mungkin dianggap sebagai satu *node* tunggal. Pada beberapa *web* yang mengandung lebih dari satu semantik dan banyak *link* hanya untuk navigasi atau iklan, akan mengakibatkan kesalahan perhitungan nilai kepentingan oleh PageRank dan pergeseran topik pada HITS[4]. Penempatan *link* pada *web* semantik juga menjadi permasalahan dalam kasus ini.

Pada algoritma *link analysis* berdasarkan pada dua asumsi, yaitu tautan yang disampaikan oleh seseorang dan halaman yang dikutip oleh halaman tertentu dan mungkin memiliki topik yang sama[4]. Pada umumnya dalam perhitungan *link analysis*, *web* dianggap sebagai unit terkecil dalam pembangunan graf *web*. Dengan demikian, setiap link pada setiap *web* akan dianggap sama, tanpa memperhitungkan letak dari *link* tersebut. Akibatnya, *link* yang terletak pada *block* yang kecil kemungkinannya untuk mendapat perhatian dari *user* akan mendapat bobot yang sama pula, dibandingkan dengan *block* yang terletak di tengah halaman *web* yang memiliki kemungkinan untuk mendapat perhatian oleh *user*. Sebagai solusinya, muncul metode *block level link analysis* yang memilah bagian *web* menjadi satuan *block*. Dengan demikian, setiap satuan *block* mendapat bobot yang berbeda, sehingga *link* yang berada pada *block* yang lebih besar dan memungkinkan untuk mendapat perhatian lebih dari *user* akan mendapatkan bobot yang lebih besar pula.

Block level link analysis merupakan sebuah metode pengembangan dari *link analysis* yang menjadikan suatu halaman *web* menjadi bagian-bagian kecil. Dengan menggunakan metode ini, kekurangan *link analysis* dapat tertutupi dengan pemberian bobot yang berbeda pada setiap *block*-nya didasarkan dari besarnya *block* dan kemungkinan *block* tersebut mendapat

perhatian *user*. *Block Level* HITS merupakan penerapan dari metode *block level link analysis* yang juga merupakan pengembangan dari algoritma HITS. *Block level link analysis* adalah metode yang membagi suatu *web* menjadi *block-block* yang merupakan unit terkecil, sehingga perhitungan *link analysis* menjadi lebih dapat dihandalkan. HITS adalah suatu algoritma yang menghitung nilai otoritas dan nilai penghubung untuk menentukan *web* yang lebih penting.

1.2 Perumusan Masalah

1. Bagaimana pengaruh jumlah dan letak *link* terhadap peringkat halaman *web* yang dituju?
2. Bagaimana pengaruh hasil algoritma *Block Level* HITS dalam banyaknya *link* pada *page*?

1.3 Batasan Masalah

1. Penelitian tugas akhir ini hanya fokus dalam pemberian bobot pada algoritma *Block Level* HITS saja.
2. Dataset berupa dokumen *off-line* berupa berkas teks hasil crawling.
3. Dataset yang digunakan diambil dari http://www.cs.toronto.edu/~tsap yang telah melalui tahap *pre-processing*.

1.4 Tujuan Penelitian

1. Menganalisis pengaruh jumlah dan letak *link* terhadap peringkat halaman *web* yang dituju.
2. Menganalisis pengaruh hasil algoritma *Block Level* HITS perbandingan jumlah *link* pada suatu *page*.

1.5 Metodologi Penyelesaian Masalah

1. Studi literatur
   Pada tahap ini akan dilakukan pembelajaran konsep teori-teori tentang *link analysis* dan algoritma HITS yang digunakan, serta informasi lainnya yang menunjang pembuatan tugas akhir ini dari berbagai macam sumber.
2. Pengumpulan data
   Pada tahap ini data yang diambil untuk dataset diolah untuk penentuan peringkat dengan menggunakan algoritma *Block Level* HITS. Data yang diambil berupa data HTML dengan *link* yang menunjuk pada suatu *web* tertentu dalam dokumen *web*.
3. Pemodelan sistem
   Tahap ini meliputi analisis kebutuhan sistem dalam perhitungan pemeringkatan berdasarkan kebutuhan yang telah diidentifikasi.
4. *Testing* dan analisis hasil
   Pada *testing* akan dilakukan percobaan kepada sistem yang selanjutnya akan diketahui sistem tersebut telah berjalan dengan baik sesuai dengan tujuan yang telah ditentukan sebelumnya atau tidak. Hasil yang didapat akan berupa pembobotan *web* yang selanjutnya akan dianalisis mengenai pengaruh letak dan jumlah *link* terhadap peringkat *web* tersebut.