

## ABSTRAKSI

Sebuah mesin pencari dituntut dapat memberikan hasil pencarian yang tepat dan benar – benar *Relevant* terhadap keinginan *user*. Konsep dari mesin pencari ini menggunakan *information retrieval*. Dalam *information retrieval*, terdapat dua jenis dokumen, yaitu *free text (unstructured document)* dan *fielded text (structured document)*. Dokumen HTML merupakan salah satu jenis *fielded text*. Dalam pencarian dokumen dalam bentuk HTML, harus diperhatikan adanya faktor tingkat kepentingan dari setiap bagian dari dokumen. Tingkat kepentingan dokumen berdasar *meta tag title, keyword, body, dan H1* dengan menggunakan metode *Pagerank* dan metode Okapi (*BM25F*) untuk menghasilkan tingkat performansi parameter-parameter yang dibutuhkan.

Metode *BM25F* diimplementasikan pada ruang lingkup pembobotan dokumen (*weighting*), sedangkan metode *Pagerank* diimplementasikan pada pembobotan node dokumen (*ranking*). Untuk mengkombinasikan kedua metode tersebut, masing-masing metode menghasilkan nilainya yang kemudian akan dikalikan untuk mengetahui nilai *similarity score* setiap dokumen yang dihitung. Dengan menerapkan kedua metode tersebut yang kemudian dikombinasikan, memungkinkan akan menghasilkan tingkat performansi dari parameter (*Precision, Recall dan Interpolated Average Precision*) yang lebih baik dari metode yang hanya berdiri sendiri misalnya hanya menggunakan metode *BM25F*.

Dari hasil pengujian pada skenario pengujian perubahan Top-N, dapat disimpulkan bahwa hasil performansi sistem yang dicapai stabil di beberapa query, hal ini terjadi karena query tersebut memiliki dokumen yang tingkat *similarity score* yang tinggi dan memiliki nilai *Pagerank* yang besar atau bisa dikatakan tingkat kepopuleran dokumennya tinggi. Pengkombinasian metode ini sangat berpengaruh terhadap nilai *Pagerank* suatu dokumen, karena jika semakin besar nilai *Pagerank* maka *similarity score* yang dihasilkan juga akan lebih tinggi dari dokumen yang lain. Akan tetapi, pengkombinasian Metode *BM25F* dan metode *Pagerank* ini memiliki kelemahan pada parameter *Recall* (kelengkapan) yang dihasilkan akan menurun hingga < 50% ketimbang menggunakan metode *BM25F* saja. Lain halnya dengan parameter *Precision* dan *IAP (INTERPOLATED AVERAGE PRECISION)* yang lebih baik dari metode *BM25F* saja yang dapat mencapai presentase 100%.

Kata Kunci : Information Retrieval, Metode *BM25F*, Metode *Pagerank*