

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. Faktor – faktor tersebut, yang selanjutnya disebut sebagai *static rank*, dapat dibedakan berdasar *tags* atau *markup* dalam dokumen HTML tersebut, seperti *title*, *text*, *inlinks*, *obj*, *type* dll.

Metode BM25F merupakan salah satu metode yang menerapkan perhitungan *IR score* dan *static rank*. Metode BM25F diimplementasikan pada ruang lingkup pembobotan dokumen (*weighting*), dimana didalam metode ini terdapat perhitungan khusus terhadap *field (tag)* tertentu yang dipengaruhi oleh *boost factor*. Performansi dari implementasi metode BM25F akan dilihat berdasarkan kecocokan dokumen dengan kata kunci (*query*) yang selanjutnya disebut sebagai nilai relevansi.

Dengan menerapkan metode BM25F diperoleh bahwa, dengan perubahan banyaknya jumlah dokumen (N) serta jumlah dokumen *relevant* dengan proporsi tetap yang digunakan, nilai performansi berdasarkan parameter *precision* dan *recall* dari *information retrieval system* menggunakan metode BM25F menghasilkan performansi yang baik, yaitu cenderung mendekati nilai yang stabil.

Dari hasil pengujian pada skenario *boost factor*, dapat disimpulkan bahwa hasil performansi sistem (*precision*, *recall* dan *IAP / interpolated average precision*) yang paling baik didapat pada saat menggunakan skenario default (*{title = 4};{H1 = 3};{anchor = 2};{span = 2};{text = 1}*;). Hal ini dikarenakan proporsi atau pemberian nilai *boost factor* yang seimbang pada skenario default, sesuai urutan tingkat kepentingan dari tiap *field* dalam dokumen berdasarkan asumsi. Tetapi, hal ini tidak terlepas dari faktor lain, yaitu keberagaman ada / tidaknya *field* yang diujikan dalam skenario di dalam masing – masing dokumen.

Kata kunci : *information retrieval, information retrieval system, BM25F*