

Abstraksi

Information Retrieval (IR) merupakan bagian dari *computer science* yang berhubungan dengan pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri. Proses dalam *Information Retrieval* dapat digambarkan sebagai sebuah proses untuk mendapatkan *relevant documents* dari *collection documents* melalui pencarian *query* yang diinputkan *user*. Dalam information retrieval terdapat dua jenis dokumen yaitu free text document (unstructured document) dan fielded text (structured document). Salah satu jenis dokumen fielded text adalah HTML. Dalam pencarian dokumen fielded text, terutama dokumen HTML harus diperhatikan adanya tingkat kepentingan dari bagian yang membangun dokumen tersebut, yang disebut dengan static rank. Static rank dibedakan berdasarkan tag, dalam dokumen HTML terdapat tag title, keywords, head, body, span, dll.

GBM merupakan salah satu metode yang dapat menerapkan perhitungan *static rank* dan *IR score*, untuk mendapatkan perhitungan *static rank* dalam tugas akhir ini menggunakan metode *Pagerank*. Performansi dari implementasi metode GBM akan dilihat berdasarkan kecocokan dokumen dengan kata kunci (*query*) yang selanjutnya disebut sebagai nilai relevansi.

Dari hasil pengujian didapatkan bahwa dengan parameter *precision*, *recall* dan IAP metode inverse kuadrat lebih unggul dibandingkan dengan metode yang lain. ($\frac{1}{x^2}$). Untuk pengujian perubahan nilai boost factor pada dokumen HTML, hasil performansi system terutama dilihat dari nilai IAP, yang lebih baik adalah ketika menggunakan *boost factor default* [Ahmad, 2011] {(keywords=4), (title=4), (head=3), (body=2), (span=2)}, dimana *field keywords* dan *field title* memiliki tingkat kepentingan yang sama dan paling tinggi dibandingkan tiga *field* lain. 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, Gravitation Based Model, precision, recall, IAP, document collection dan query.*