

1. Pendahuluan

1.1. Latar Belakang

Dalam era perkembangan teknologi informasi saat ini, keberadaan sebuah *search engine* telah menjadi sebuah bagian dari kehidupan sehari-hari. *Search engine* dapat membantu pengguna internet untuk menemukan informasi di Internet, dengan terlebih dahulu memasukkan sebuah kata kunci kedalam *search engine* tersebut [5].

Kelemahan dari *search engine* adalah, *search engine* hanya akan mencari informasi yang dicari dengan tepat, dengan kata kunci atau disebut juga dengan *query* yang tepat pula. Disinilah terjadi masalah pada proses pencarian informasi melalui *search engine*. User biasanya menginputkan *query* yang terlalu panjang atau pendek dan tidak spesifik. *Query* yang dimasukkan biasanya bermakna sangat luas dan tidak menggambarkan secara spesifik informasi yang dicari. Untuk itulah dalam Information Retrieval, diperlukan sebuah *Relevance Feedback*. *Relevance Feedback* adalah sebuah teknik, dimana kata kunci (*Query Awal*) akan diformulasi ulang untuk menghasilkan *query* yang lebih baik, berdasarkan dari dokumen yang telah berhasil diretrieve dari *query* awal [5].

Dalam Tugas Akhir ini, akan dibahas tentang Information Retrieval dengan menggunakan *Blind Relevance Feedback*. *Blind Relevance Feedback* merupakan salah satu metode *Relevance Feedback* dalam Information Retrieval, dimana user tidak perlu memberikan *feedback* secara aktif kepada sistem, tetapi sistem akan melakukan *Query Expansion* dengan menganggap bahwa dokumen yang mempunyai ranking tinggi adalah dokumen yang relevan terhadap pencarian user. Alasan dipilihnya metode *Blind Relevance Feedback* adalah karena, dibandingkan dengan metode *Relevance Feedback* yang lain, *Blind Relevance Feedback* mempunyai keunggulan mengotomasi bagian manual dari *Relevance Feedback*, dan tidak memerlukan assessor, sehingga hasil *Relevance Feedback* tidak akan terpengaruh faktor Assesor. Teknik *Blind Relevance Feedback* yang digunakan pada tugas akhir ini adalah *Blind Relevance Feedback* Type 1 dan 2, dimana perbedaan dari BRF (*Blind Relevance Feedback*) tipe 1 dan 2 adalah *Query Expansion Collection* yang digunakan [2].

1.2. Perumusan Masalah

Perumusan masalah dalam tugas akhir ini adalah sebagai berikut :

1. Bagaimana perbandingan performa *Information Retrieval* menggunakan masing-masing tipe *Blind Relevance Feedback*, dan Kombinasi dari kedua tipe *Blind Relevance Feedback* dalam meretrieve dokumen yang dicari?
2. Bagaimana mengetahui tipe BRF mana yang tepat untuk digunakan, apakah tipe BRF yang berbeda akan mengembalikan hasil yang berbeda?
3. Bagaimana menentukan saat yang tepat untuk menggunakan BRF tipe 1 atau tipe 2, dan menentukan saat yang tepat untuk tidak menggunakan BRF ?

Adapun batasan masalah dalam Tugas Akhir ini adalah :

1. *Relevance Feedback* yang digunakan hanya *Blind Relevance Feedback*.
2. Dataset yang digunakan adalah TREC WT10g [WTX010~WTX019] Berjumlah 500 Files, dimana *data collection* yang didapatkan adalah *data*

collection dari hasil crawl situs di Internet. Sedangkan *Query* uji yang digunakan adalah *query* dari TREC web track.

3. *Query Expansion Collection* yang digunakan dalam *Blind Relevance Feedback* Type 2 adalah WordNet.
4. Tugas Akhir ini berfokus pada *Blind Relevance Feedback*, sehingga dalam Tugas Akhir ini tidak akan membahas secara mendalam proses indexing, dan proses *preprocessing* lainnya sebelum proses *Relevance Feedback*.

1.3. Tujuan

Secara umum, Tujuan dari pembuatan Tugas Akhir ini adalah :

1. Untuk mengukur dan menganalisis sejauh mana peningkatan hasil *retrieve* dokumen yang dihasilkan oleh *Blind Relevance Feedback* Type 1 dan 2, jika dibandingkan dengan hasil *retrieve* dokumen tanpa menggunakan *Blind Relevance Feedback*.
2. Untuk menganalisis, tipe BRF mana yang tepat untuk digunakan dalam Topik Uji Studi Kasus Tugas Akhir ini. Jika ada perbedaan performa antara BRF tipe 1 dan BRF tipe 2, maka hasil analisa juga diharapkan dapat menunjukkan faktor-faktor yang mempengaruhi perbedaan performa BRF tersebut.
3. Untuk menganalisis, kapan saat yang tepat untuk menggunakan BRF tipe 1, atau Tipe 2. Dan kapan saat untuk tidak menggunakan BRF sama sekali.
4. Secara khusus, diharapkan setelah Tugas akhir ini selesai, dapat dibuktikan bahwa *Information Retrieval* yang menggunakan *Blind Relevance Feedback*, menghasilkan *Information Retrieval* yang lebih baik dalam kemampuannya meretrieve dokumen yang relevan daripada sebelum menggunakan *Blind Relevance Feedback*.

Hipotesa awal untuk Tugas Akhir ini adalah, *Information Retrieval* dengan menggunakan *Blind Relevance Feedback* , akan mengembalikan atau meretrieve dokumen yang lebih relevan bagi user dibandingkan dengan *Information Retrieval* yang tidak menggunakan *Blind Relevance Feedback*, terlepas dari tipe dari *Blind Relevance Feedback* yang digunakan.

1.4. Metodologi Penyelesaian Masalah

Metodologi yang digunakan dalam menyelesaikan Tugas Akhir ini adalah :

1. Studi Literatur
Mempelajari landasan teori dari referensi-referensi yang ada tentang *Information Retrieval*, *Relevance Feedback*, *Blind Relevance Feedback*, dan literatur-literatur lainnya yang membantu baik itu dalam penelitian maupun dalam pembuatan aplikasi.
2. Pembangunan perangkat lunak
 - a) Analisis dan Perancangan
Melakukan analisis dan perancangan perangkat lunak dengan menggunakan metode *Unified Modelling Language* (UML)
 - b) Pengkodean
Implementasi dari hasil perancangan ke dalam pemrograman komputer dengan menggunakan teknik pemrograman berorientasi objek.
 - c) Pengujian
Adapun skenario pengujian yang akan dilakukan adalah sebagai berikut:
 - 1) Melakukan penyimpanan *data collection*.

- 2) Melakukan indexing terhadap informasi-informasi yang terdapat dalam *data collection*.
- 3) Melakukan proses pencarian terhadap *data collection* dengan menggunakan *query* uji yang telah ditentukan. Pengujian pada tahap ini dilakukan terhadap *Information Retrieval* tanpa *Blind Relevance Feedback*, dan *Information Retrieval* dengan menggunakan *Blind Relevance Feedback*.
- 4) *Query* uji sebisa mungkin akan dibuat untuk tidak memihak, dalam artian *query* uji tidak akan membuat sebuah IR lebih baik dari IR yang lainnya.
- 5) Melakukan analisis dari hasil pengujian aplikasi.

3. Analisis Hasil

Pada tahap ini, akan dilakukan analisis hasil *query expansion* dengan menggunakan beberapa sampel kasus sebagai berikut :

- a. *Query* uji dimasukkan pada sistem tanpa menggunakan *Blind Relevance Feedback*.
- b. *Query* uji dimasukkan pada sistem dengan menggunakan *Blind Relevance Feedback*.

Parameter yang digunakan dalam pengujian sistem ini adalah :

- 1) *Precision*
Rasio dari seberapa banyak dokumen yang ter-*retrieve*, relevan terhadap user.
- 2) *Recall*
Rasio dari seberapa banyak dokumen relevan yang seharusnya ter-*retrieve*, berhasil di *retrieve*.
- 3) MAP (*Mean Average Precision*)
Mean dari *Average Precision* setelah dokumen yang relevan berhasil di *retrieve*. Nilai MAP didapat dari *Average precision* dibagi jumlah set *Query* uji yang digunakan.
- 4) nDCG
- 5) R-*Precision*
Sebuah pengukuran yang mengkomplemenkan MAP. Menggambarkan keseluruhan *Recall* dari *Top Ranked Document*.

4. Pengambilan Kesimpulan dan Pembuatan Laporan

Menarik kesimpulan dari hasil analisis yang telah dilakukan terhadap sistem, serta mendokumentasikan hasil perancangan, implementasi, pengujian, dan analisis ke dalam suatu laporan.