1. PENDAHULUAN

1.1 Latar belakang masalah

Perkembangan teknologi di dunia Internet semakin pesat. Dunia Internet sekarang ini memang sulit untuk dipisahkan dari dunia nyata. Sebagian besar aktivitas *user* dilakukan melalui Internet dan kadang *user* membutuhkan bantuan di dalam dunia Internet untuk mencari referensi bagi pekerjaannya. Maka dari itu, dibuat *search engine* untuk membantu *user* sehingga mempermudah *user* untuk mendapatkan bahan referensi bagi pekerjaannya.

Oleh karena itu, *search engine* dituntut agar dapat memberikan hasil pencarian yang tepat dan akurat sesuai dengan keinginan *user*. Akan tetapi, hasil pencarian yang didapat kadang tidak relevan dengan yang diinginkan *user*. *Search engine* sendiri menggunakan *Information Retrieval* (IR) sebagai konsep dasarnya dimana *Information retrieval* merupakan suatu sistem yang digunakan untuk menemukan kembali (*retrieve*) informasi-informasi yang relevan terhadap kebutuhan *user* dari suatu kumpulan informasi berdasarkan *query* dari *user* [11]. Sehingga dibutuhkan solusi agar meningkatkan performansi dari *Information Retrieval* yaitu dengan *query reformulation techniques* (teknik memodifikasi *query* yang ada).

Query reformulation techniques sendiri terdiri dari 2 yaitu *query expansion methods* dan *query reweighting methods* [1]. Untuk kasus Tugas Akhir ini adalah *query expansion method* dimana *query* awal yang diinputkan oleh *user* diperluas dengan *keyword* baru. Di dalam tugas akhir ini sendiri, akan dikembangkan suatu metode *Query Expansion* (QE) yang baru berdasarkan penarikan kesimpulan dari *fuzzy rules* dan *pseudo relevance feedback*. Metode ini menggunakan fungsi-fungsi keanggotaan untuk merepresentasikan atribut dari *term* yang digunakan (dalam kasus ini digunakan *Relevant Frequency*, *inverse document frequency*, dan *Relevance Degree*) dan mengkalkulasikan fungsi-fungsi keanggotaan tersebut menggunakan *fuzzy rules* untuk mendapatkan nilai *Relevance Degree* dari calon *additional keywords* dan menggabungkan calon *additional keywords* yang memiliki nilai *Relevance Degree* paling besar dengan *query* awal inputan dari *user*. Model dengan perluasan metode ini diharapkan mendapatkan nilai *Interpolated Average Precision* (IAP) yang lebih baik dibandingkan model awal yang menggunakan *Vector Space Model* (VSM).

1.2 Perumusan masalah

Berdasarkan uraian diatas maka permasalahan yang muncul dan yang menjadi objek penelitian pada Tugas Akhir ini :

1. Bagaimana proses pencarian calon *additional keywords* dari *relevant documents* yang ada menggunakan *pseudo relevance feedback*.

2. Bagaimana proses pencarian nilai *Relevance Degree* dengan menggunakan *QE+fuzzy rules*.
3. Bagaimana perbandingan nilai IAP antara model awal yang diperluas dengan *QE+fuzzy rules* dengan model awal yang menggunakan vector space model

Batasan masalah agar tidak meluasnya materi pembahasan dalam tugas akhir ini adalah sebagai berikut :

1. Analisis data dilakukan terhadap standar koleksi dokumen untuk *information retrieval* yang didapat dari <ftp://ftp.cs.cornel.edu/pub/smart/med/> yang bertipe file '.txt' dimana sudah terdapat kumpulan *relevance judgement*.
2. *Query* yang diinputkan akan ditentukan dan menggunakan *query* khusus untuk pengujian dari <ftp://ftp.cs.cornel.edu/pub/smart/med/>.
3. Aplikasi hanya melakukan *word indexing* dan tidak melakukan *phrase indexing*.
4. Algoritma *stemming* yang digunakan menggunakan algoritma *Potter*.
5. *Query* yang diuji oleh aplikasi hanya bisa berbentuk *simple query* atau tidak bisa menggunakan *Boolean operation* atau *operation* yang lain.
6. Banyak dokumen dan calon *additional keywords* yang akan digunakan hasil dari *pseudo relevance feedback* hanya diambil top 5 dokumen teratas (dengan acuan nilai *similarity* terbesar) dan top 4 kata sebagai calon *additional keywords* (dengan acuan bobot *term* terbesar).
7. Koleksi dokumen dan *query* yang digunakan menggunakan bahasa inggris.
8. Pengujian performansi akan dilakukan dengan membandingkan nilai *Precision, Recall*, dan IAP yang didapat hanya melalui model awal yang diperluas dengan *QE+fuzzy rules* dengan model awal yang menggunakan vector space model.

1.3 Tujuan

Secara umum tujuan penulisan yang ingin dicapai dalam tugas akhir ini adalah:

1. Merancang dan membangun suatu *Information Retrieval System* berupa *search engine* yang menggunakan VSM dengan pengembangan *QE+fuzzy rules*.
2. Menganalisa nilai *Precision, Recall*, dan IAP yang didapat dan membuktikan apakah VSM dengan *QE+fuzzy rules* mampu memberikan hasil yang memuaskan bila dibandingkan dengan VSM biasa.

1.4 Metodologi penyelesaian masalah

Metodologi yang digunakan untuk menyelesaikan masalah dalam Tugas Akhir ini adalah :

1. Studi Literatur
Studi literatur dari beberapa buku, jurnal, artikel yang membahas tentang *Information Retrieval, Query Expansion, Fuzzy Logic*

2. Analisis dan Desain
Tahap ini meliputi analisis kebutuhan serta penyelesaian masalah untuk merancang perangkat lunak *search engine* dengan VSM yang dikembangkan dengan *QE+fuzzy rules*.
3. Implementasi Sistem
Tahap ini meliputi pembangunan perangkat lunak yang telah dirancang pada tahap sebelumnya. Pembangunan perangkat lunak lebih ke arah *web-based* dengan menggunakan PHP dan database MySQL.
4. Analisis dan Pengujian
Pada tahapan ini yang dilakukan adalah melakukan pengujian terhadap perangkat lunak yang dibangun dan sekaligus melakukan analisis terhadap hasil pemrosesan perangkat lunak. Analisa performansi dari *search engine* ini setelah digunakan *Query Expansion* akan dinilai dari IAP yang dihasilkan dari model ini dan kemudian akan dibandingkan dengan nilai IAP dari model awalnya.
5. Penyusunan dan Laporan
Hasil penelitian akan disusun menjadi suatu laporan yang meliputi aspek-aspek dalam penelitian yaitu teori, perancangan dan implementasinya, serta membuat kesimpulan dari hasil penelitian tersebut.

1.5 Sistematika Penulisan

Sistematika Penulisan Tugas Akhir ini terdiri dari 5 bab yaitu :

- BAB I Pendahuluan**
Bab ini membahas kerangka penelitian dalam tugas akhir, meliputi latar belakang, perumusan masalah, batasan masalah, tujuan perancangan dan metodologi yang digunakan dalam perancangan sistem.
- BAB II Landasan Teori**
Bab ini menjelaskan seluruh teori yang menjadi landasan konseptual dan mendukung penyelesaian tugas akhir ini.
- BAB III Analisis dan Perancangan Sistem**
Bab ini membahas mengenai pengumpulan data analisis dan perancangan perangkat lunak yang terdiri dari perancangan struktur data , perancangan modul dan *interface*.
- BAB IV Implementasi dan Pengujian Sistem**
Bab ini membahas implementasi detail sistem dan pengujian terhadap sistem
- BAB V Kesimpulan dan Saran**
Berisi tentang kesimpulan dan saran yang dapat diambil dari keseluruhan sistem yang telah dibuat.