

1. PENDAHULUAN

1.1 Latar Belakang

Dampak dari perkembangan dunia teknologi informasi membuat banyak orang berlomba-lomba untuk menggali informasi yang semakin banyak tersebarluaskan. Melihat hal itu, saat ini tentunya dibutuhkan suatu mekanisme tertentu dalam proses pencarian informasi untuk mendapatkan hasil pencarian informasi yang relevan. Pencarian informasi dalam sebuah dokumen mungkin saja bisa dilakukan dengan cara manual. Namun, tentunya hal itu akan membutuhkan waktu yang sangat banyak jika dilakukan dalam lingkup dokumen yang besar. Solusi yang bisa diterapkan untuk mengatasi masalah tersebut adalah dengan suatu *Information Retrieval System*. *Information Retrieval System* membuat *user* dapat memperoleh suatu dokumen dari kumpulan dokumen yang ada berbasiskan kata kunci.[11] *Information Retrieval System* mampu mencari secara efektif dalam jumlah data yang sangat banyak karena suatu *Information Retrieval System* membangun indeks dari dokumen – dokumen yang ada, sehingga ketepatan antara kata kunci dengan dokumen yang dicari dapat tercapai.

Term weighting merupakan tahapan dalam *indexing*. *Term weighting* adalah proses pembobotan setiap *term* dalam dokumen. Hal ini diperlukan karena dalam suatu *Information Retrieval System* kuncinya adalah menghitung similaritas dari suatu *query* dengan dokumen.[7] Nantinya *term* tersebut berkaitan dengan similaritas dari *query* dan dokumen. Semakin besar nilai similaritasnya, semakin relevan pula antara *query* dan dokumen.

Dalam Tugas Akhir ini, diterapkan suatu metode dalam pengembangan *term weighting* yang merupakan salah satu model dari *Windows Based Model*, yaitu *Core Windows Based*. Dalam *Windows Based Model* digunakan suatu pendekatan dengan dua ide atau gagasan baru.[7] Ide pertama adalah *query* yang letaknya berdekatan dalam suatu dokumen mempunyai nilai similaritas yang lebih tinggi. Sedangkan ide kedua adalah dalam suatu *query* ada *term* yang lebih penting dari lainnya yang disebut *core words*. [7] Dalam *Core Windows Based*, dikelompokkan antara kata yang penting (*core words*) dengan kata yang kurang penting (*surrounding words*). [7] Kemudian setiap *windows* yang mempunyai *core words* di dalamnya disebut *active windows*. [7] Sedangkan setiap *windows* yang mempunyai *surrounding words* disebut *common windows*. Karena suatu *active windows* mengandung *core word*, maka *active windows* mempunyai bobot yang lebih besar dari *common windows*. [7] Pemilihan *core word* dalam tugas akhir ini menggunakan metode *TF-IDF* dan *IDF*. Untuk mengukur performansinya digunakan parameter *Precision*, *Recall*, *IAP* dan *F-Measure*.

Dengan menerapkan metode *Core Windows Based* dalam *Term weighting* pada suatu *Information Retrieval System*, diharapkan tingkat relevansi *query* dengan dokumen menjadi lebih tinggi. Sehingga tingkat efektifitas dalam pencarian dokumen menjadi lebih baik. Hal itu bisa dilihat dengan memperhatikan nilai *precision* dan *recall*.

1.2 Perumusan Masalah

Rumusan masalah pada Tugas Akhir ini adalah :

- a. Bagaimana pengaruh penentuan ukuran dari *width windows* dalam *Core Windows Based* untuk memperoleh hasil yang optimal
- b. Bagaimana dampak metode pemilihan *core word* terhadap performansi metode *Core Windows Based* yang dihitung dengan menggunakan parameter *Precision Recall* dan IAP.

1.3 Tujuan

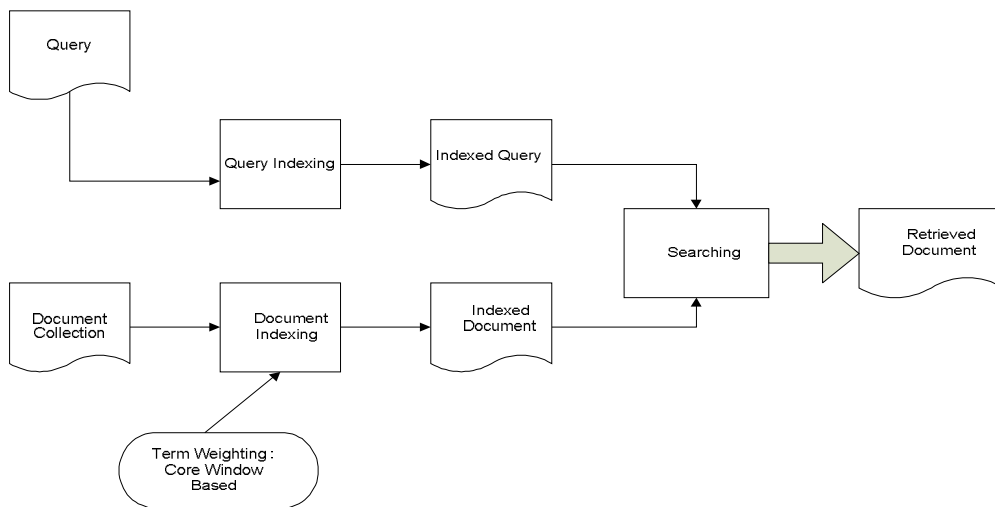
Tujuan dari Tugas Akhir ini adalah :

- a. Mengimplementasikan penggunaan *Term weighting* metode *Core Windows Based* pada *Information Retrieval System*.
- b. Menguji performansi *Core Windows Based* dengan menganalisa pengaruh perubahan *width windows* pada metode *Core Windows Based*.
- c. Menguji performansi *Core Windows Based* dengan pemilihan *core word* menggunakan *TF-IDF* dan *IDF*.

1.4 Batasan Masalah

- a. Tugas Akhir ini menitik beratkan pada pengukuran performansi dari metode *Core Windows Based* pada *Information Retrieval System* menggunakan *Precision*, *Recall*, IAP, dan *f-measure*.
- b. Pemilihan *Core word* diimplementasikan dengan menggunakan metode *TF-IDF* dan *IDF* dengan *threshold* yang ditentukan.
- c. *Dataset* yang digunakan dalam Tugas Akhir ini yang berupa kumpulan dokumen dan *query* menggunakan bahasa Inggris. *Dataset* yang digunakan berbentuk *.txt dan dokumen yang digunakan merupakan dokumen yang termasuk *unstructured document*. *Dataset* yang digunakan berasal dari ftp.cs.cornell.edu/pub/smart . Di dalam *dataset* yang digunakan terdapat kumpulan dokumen dan *query* nya beserta *relevance judgement* untuk tiap-tiap dokumen dan berisi data index untuk masing-masing dokumen dan *query*.
- d. *Preprocessing* dalam implementasi Tugas Akhir ini, menggunakan *data index* yang didapatkan dari *dataset*.

1.5 Deskripsi Sistem



Gambar 1 – 1 : Deskripsi Sistem

Input :

- *Query words* yang sudah ditentukan

Output :

- Dokumen yang sesuai dengan *query* yang ditentukan
- Nilai performansi diukur dengan *precision*, *recall*, dan IAP

1.6 Metodologi Penyelesaian Masalah

a. Studi Literatur

Pengumpulan semua referensi dari berbagai sumber yang berkaitan dengan penggunaan pengembangan *Term weighting* metode *Core Windows Based* pada *Information Retrieval System*

b. Pengumpulan Data

Mencari dan menentukan *dataset* yang digunakan dalam implementasi metode *Core Windows Based* pada *Information Retrieval System*.

c. Pemodelan Sistem

- Perancangan : Dilakukan perancangan sistem dalam proses *Term weighting* dengan menggunakan metode *Core Windows Based* pada *dataset* yang telah ditentukan
- Implementasi : Menerapkan proses pengembangan *term weighting* dengan menggunakan metode *Core Windows Based* dengan menggunakan data uji berupa *dataset* yang telah ditentukan di awal.
- Testing : Melakukan suatu pengujian terhadap sistem yang dibangun untuk melihat apakah sistem tersebut sudah sesuai harapan

d. Analisa dan Evaluasi Hasil

- Untuk menganalisis performansi *Core Windows Based*, inputkan *query* ke dalam sistem lalu lihat nilai *similarity* yang didapatkan terhadap dokumen kemudian hitung nilai *precision*, *recall* dan IAP.
- Ubah *width windows* dengan nilai yang berbeda untuk mengetahui performansinya