

## 8. 1. Pendahuluan

### 1.2 1.1 Latar Belakang Masalah.

Pada saat ini dimana jumlah dokumen yang tersedia sangat besar, pencarian secara manual dapat dilakukan dengan membaca setiap dokumen pada koleksi dokumen untuk mendapatkan dokumen yang tepat dan sesuai kebutuhan. Namun, pencarian seperti itu membutuhkan waktu yang lama jika jumlah dokumen sangat banyak. Salah satu solusi untuk mengatasi masalah tersebut adalah dengan menggunakan *Information Retrieval System*. Proses dalam *Information Retrieval* dapat digambarkan sebagai sebuah proses untuk mendapatkan *relevant documents* dari *collection documents* yang ada melalui pencarian *query* yang diinputkan *user*[2]. Sedangkan *query* dalam *Information Retrieval* merupakan kata kunci berupa satu atau beberapa term yang akan dicari. *Information retrieval systems* yang baik adalah sistem yang mampu mengambil dokumen relevan dan kemudian mengurutkan dokumen relevan tersebut pada urutan teratas[2]. Agar diperoleh sistem yang dapat mengambil dokumen yang relevan dan mengurutkan pada urutan teratas dibutuhkan suatu metode *matching* suatu *query* dengan koleksi dokumen menggunakan metode yang optimal dalam menentukan nilai *similarity score* suatu dokumen.

Metode yang digunakan dalam proses penghitungan nilai *similarity score* saat ini sebagian besar hanya memperhitungkan kemunculan suatu *term* dalam dokumen namun tidak memperhitungkan jarak antara term yang muncul dalam dokumen. Padahal jarak antar *term query* didalam dokumen sangat menentukan tingkat *similarity* antara dokumen dengan *query* terutama bila term tersebut berbentuk frasa. Jarak *term* satu dengan *term query* yang lain yang jauh dapat dikatakan tidak bisa menunjukkan arti yang jelas. Biasanya di dalam *query* terdiri dari beberapa *term query*, dimana semua *term query* tersebut menunjukkan sebuah arti[11]. Misal terdapat *query "brain cancer dangerous"* dan dalam dokumen 1 terdapat *term "brain cancer"* dalam dokumen 2 terdapat *term "cancer doesn't have negative effect to brain"* maka dari contoh diatas didapatkan dokumen 1 dengan jarak antar *term* yang kecil lebih relevan dibanding dengan dokumen 2 yang mempunyai jarak antar *term* lebih besar. *Term* yang muncul dengan jarak berdekatan dalam dokumen berkontribusi lebih besar terhadap nilai *similarity* dibandingkan dengan *term* yang muncul dengan jarak yang berjauhan. Semakin dekat jarak suatu *term* dalam dokumen maka semakin besar nilai *similarity* antara *query* dan dokumen[11].

Oleh sebab itu dalam Tugas Akhir ini akan diimplementasikan dan dianalisis performansi dari penerapan metode *Dynamic Window Based Method* dalam *IR System*.

Dalam menganalisa hasil penerapan dari metode *Dynamic Window Based* akan digunakan metode yang lain yaitu Probabilistik sebagai pembanding.

### 1.3 1.2 Perumusan Masalah.

Rumusan masalah dari tugas akhir ini adalah :

- a) Bagaimana menerapkan metode *Dynamic Window Based* dalam *information retrieval systems*? Dalam menganalisis performansi nya akan digunakan metode Probabilistik sebagai pembanding.

- b) Bagaimanakah perbandingan performansi sistem dengan metode *Dynamic Window Based* dibandingkan dengan metode Probabilistik?
- c) Bagaimana pengaruh ukuran lebar *window* terhadap performansi *information retrieval system*.

#### 1.4 1.3 Batasan Masalah.

Adapun batasan masalah dari tugas akhir ini adalah :

- a) Penelitian pada Tugas Akhir ini berfokus pada pengimplementasian metode *Dynamic Window Based* pada IRS serta analisis performansinya dengan menggunakan parameter *Precision, Recall*, serta *IAP*.
- b) Batasan lebar *window* yang digunakan adalah lebih besar dari 1 dan kurang dari sama dengan 20, karena rata-rata panjang *query* yang ada tidak melebihi 20 *term*.
- c) Mengasumsikan bahwa hasil dari proses *preprocessing* yang digunakan sudah benar.
- d) Perhitungan nilai performansi sistem dibatasi/dihitung hanya pada 20 dokumen yang ter-*retrieved* pertama, karena diharapkan performansi sistem akan lebih terlihat untuk masing-masing metode.
- e) Koleksi dokumen yang digunakan dalam tugas akhir ini yang berupa kumpulan dokumen dan *query* menggunakan bahasa Inggris. Koleksi dokumen yang digunakan berbentuk \*.txt dan dokumen yang digunakan merupakan dokumen yang termasuk *unstructured document*. Dalam Tugas Akhir ini digunakan koleksi dokumen yang berasal dari [ftp.cs.cornell.edu/pub/smart](http://ftp.cs.cornell.edu/pub/smart) . Di dalam koleksi dokumen yang digunakan terdapat kumpulan dokumen dan *query* nya beserta *relevance judgment* untuk tiap-tiap dokumen. *Query* yang digunakan merupakan *simple query*.

#### 1.5 1.4 Tujuan.

Tujuan dari tugas akhir ini adalah :

- a) Mengimplementasikan metode *Dynamic Window Based* pada *information retrieval systems*.
- b) Menganalisis performansi penggunaan metode *Dynamic Window Based* pada *information retrieval system*. Dalam menganalisa performansi hasil penerapan dari metode *Dynamic Window Based* akan digunakan metode yang lain yaitu metode Probabilistik sebagai pembanding.
- c) Menganalisis pengaruh ukuran lebar *window* terhadap performansi *information retrieval systems* yang diukur dengan menggunakan parameter F-Measure dan IAP.

## 1.6 1.5 Metodologi Penyelesaian Masalah.

### 1.7 Metodologi penyelesaian masalah yang akan digunakan adalah :

a. Studi literatur

Merupakan tahapan dalam mempelajari konsep dan teori pendukung untuk memecahkan permasalahan. Pencarian sumber dan referensi berupa buku, makalah, jurnal dan media yang lain seperti internet yang berhubungan dengan konsep *information retrieval*, dan metode *Dynamic Window Based*, serta informasi lainnya yang menunjang pembuatan tugas akhir ini.

b. Pengumpulan data

Pengumpulan dokumen yang akan dijadikan sebagai data set untuk analisis dan pengimplementasian metode *Dynamic Window Based* dalam *information retrieval system*.

c. Perancangan Sistem.

Mempersiapkan data set yang akan digunakan dalam proses *matching* menggunakan metode *Dynamic Window Based*. Kemudian dilakukan perancangan sistem yang nantinya akan dilakukan proses implementasi metode *Dynamic Window Based* pada IRS.

d. Implementasi.

Tahapan implementasi ini akan dilakukan proses pengimplementasian metode *Dynamic Window Based* pada IRS. Di tahap ini terdiri dari :

- Dokumen uji yang telah disiapkan sebelum dilakukan proses perhitungan nilai *similarity score* yaitu dilakukan proses *tokenisasi, stopword removal, stemming, term weighting* terlebih dulu.
- Menerapkan metode *Dynamic Window Based* dalam IRS untuk perhitungan nilai *similarity score* dengan mengaplikasikannya kedalam kode program.

e. Testing dan analisis Hasil.

Dilakukan proses pengujian terhadap sistem yang sudah dibangun apakah sistem sudah sesuai dengan yang diharapkan.

- Untuk menganalisis performansi metode *Dynamic Window Based* diinputkan suatu *query* ke dalam sistem kemudian nilai *similarity* terhadap dokumen yang didapatkan dihitung nilai *precision, recall, dan IAP* nya. Kemudian dilakukan prosedur yang sama dengan lebar *window* yang berbeda, untuk mengetahui pengaruh lebar *window* terhadap performansi sistem. Hasil pengujian dianalisis berdasarkan parameter *precision, recall, dan IAP*. Dari hasil analisis tersebut diambil kesimpulan mengenai penerapan metode *Dynamic Window Based* pada *IR System*.

## 1.8 1.6 Sistematika Penulisan.

Sistematika Penulisan Tugas Akhir ini terdiri dari lima bab dengan disertai lampiran terkait pelaksanaan tugas akhir 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	Dasar 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.
BAB IV	Implementasi, Pengujian dan Analisis	Bab ini membahas implementasi detail sistem dan pengujian terhadap sistem.
BAB V	Kesimpulan dan Saran	Berisi tentang kesimpulan dan saran sebagai hasil dari analisis dan implementasi Tugas Akhir.