

# 1. PENDAHULUAN

## 1.1 Latar Belakang

Menyusun ulang query (*query reformulation*) yang dimasukkan oleh user adalah hal yang sering dilakukan dalam *information retrieval*. Hal ini dilakukan untuk mengatasi ketidaksesuaian antara query yang dimasukkan oleh user dengan informasi yang ingin didapatkannya. *Query reformulation* yang sering dipakai adalah dengan *Query Expansion*, yaitu dengan memanjangkan query yang dimasukkan oleh user dengan menambahkan beberapa term kedalamnya. Query yang dimasukkan oleh user pada umumnya pendek dan *query expansion* dapat melengkapkan informasi yang ingin dicari user.

Dalam temu kembali dokumen beberapa pendekatan yang dapat digunakan dalam meranking term yang akan digunakan dalam *query expansion*. Dua diantaranya adalah melalui pendekatan *Cooccurrence* dan *Probabilistic*. Pendekatan pertama merupakan pendekatan yang berbasis kepada pengukuran kemunculan kandidat term dan term dari query pada document yang didapatkan atau yang sering disebut sebagai *cooccurrence approach*. Yang kedua adalah dengan pendekatan probalistik (*probabilistic approach*), pendekatan ini berbasis kepada distribusi sebuah term dalam *document collection* dan hasil pencarian teratas.

Pendekatan pertama menghasilkan term-term yang sering muncul bersamaan dengan term yang terdapat pada query user. Sedangkan untuk pendekatan kedua, *query expansion* menghasilkan term-term yang memiliki tingkat kepentingan yang tinggi. Karena kedua pendekatan ini menghasilkan term-term yang berbeda, maka sangat mungkin dilakukan penggabungan untuk mendapatkan hasil dari *query expansion* yang lebih baik.

## 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 term* yang akan dipakai dalam memanjangkan *query* dengan menggunakan metode *Co-Occurence* dan *Probabilistic*.
2. Bagaimana proses penggabungan metode *Co-Occurence* dan *Probabilistic* dalam pencarian calon *additional term*.
3. Bagaimana perbandingan performansi pencarian antara metode pada Tugas Akhir ini dengan *original matching function*-nya.
4. Bagaimana pengaruh performansi hasil penggabungan metode *Co-Occurence* dan *Probabilistic* terhadap masing-masing metode.

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

1. Koleksi dokumen dan kata kunci yang digunakan adalah dokumen dalam teks bahasa Inggris.
2. Koleksi dokumen menggunakan koleksi dokumen yang diambil dari <ftp://ftp.cs.cornell.edu/SMART>. Koleksi dokumen yang dipakai adalah

koleksi dokumen dengan kategori MED yang terdiri dari 1033 buah dokumen.

3. Pada koleksi dokumen tersebut sudah terdapat kumpulan kata kunci yang berjumlah sepuluh buah beserta *relevance judgement* untuk setiap kata kunci(*Query*) tersebut.
4. Stopword menggunakan daftar kata yang terdapat pada koleksi dokumen(yang diambil dari <ftp://ftp.cs.cornell.edu/SMART>).
5. Model sistem temu kembali informasi yang digunakan pada awal pencarian adalah model ruang vektor(Vector Space Model), dimana dokumen maupun kata kunci direpresentasikan sebagai vektor berdimensi  $n$ , dengan  $n$  adalah jumlah kata atau *term* pada kata kunci. *Relevance feedback* pada model ruang vektor dapat dijelaskan sebagai penggeseran vektor kata kunci mendekati vektor dokumen relevan dan menjauhi vektor dokumen tidak relevan[2].
6. Jumlah dokumen teratas yang akan diproses untuk mendapatkan calon *additional term* adalah lima(5) buah dokumen(peringkat 1 sampai dengan 5).
7. Jumlah *additional term* yang digunakan untuk setiap pencarian berjumlah lima(5) buah term.
8. Pengujian menggunakan parameter IAP(*Interpolated Average Precision*).

### 1.3 Tujuan

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

1. Merancang dan membangun suatu *Information Retrieval System* yang menggunakan *Query Expansion* dengan menggabungkan Metode *Co-Occurrence* dan *Probabilistic*.
2. Menganalisis proses dari pencarian *Additional Term* dengan menggabungkan Metoda *Co-Occurrence* dan *Probabilistic* yang kemudian digunakan untuk membuktikan apakah metode ini mampu memberikan nilai IAP yang lebih baik dari metode awalnya(*Vector Space Model*).

### 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*, *Co-Occurrence Method*, dan *Probabilistic Method*.
2. Analisis dan Desain  
Tahap ini meliputi analisis kebutuhan serta penyelesaian masalah untuk merancang perangkat lunak *search engine* dengan metoda *Query Expansion* dengan menggabungkan *Cooccurrence* dan *Probalistic Method*. Desain perangkat lunak yang akan dibangun berdasarkan proses berikut :

### 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 bahasa pemrograman PHP dan database MySQL. Implementasi penggabungan kedua metoda ini dilakukan secara seri seperti yang terlihat di gambar 2.

### 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 nilai IAP yang dihasilkan dari metode ini dan kemudian akan dibandingkan dengan nilai IAP dari *original matching function*-nya.

### 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 system.

#### **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.