

# 1. PENDAHULUAN

## 1.1 Latar Belakang

Pada saat ini, kebutuhan setiap individu terhadap Internet terus meningkat. Hal ini dapat disebabkan karena semakin banyaknya fasilitas yang ditawarkan dari dunia Internet itu sendiri. Salah satu yang penting dari Internet bahwa banyaknya web-web atau blog-blog yang menyediakan beragam informasi baik berupa berita ataupun artikel tentang suatu hal. Sebagian besar aktivitas *user* dalam dunia Internet digunakan untuk mencari suatu data atau berita yang diinginkannya dari web-web atau blog-blog tersebut. Dan mayoritas dari mereka membutuhkan *search engine* sebagai acuan pencarian bahan referensi yang dicarinya. Oleh karena itu, sebuah *search engine* dituntut untuk dapat memberikan hasil pencarian yang tepat dan benar-benar *relevant* terhadap keinginan *user*. *Search engine* sendiri menggunakan *Information Retrieval* (IR) sebagai konsep dasarnya.

*Information Retrieval* (IR) merupakan bagian dari *computer science* yang berhubungan dengan pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri. Proses dalam *Information Retrieval* dapat digambarkan sebagai sebuah proses untuk mendapatkan *relevant documents* dari *collection documents* melalui pencarian *query* yang diinputkan *user*. Ada tiga paradigma dalam penelitian di *Information Retrieval* (IR) yaitu *Probabilistic IR*, *Knowledge-based IR*, dan *Artificial Intelligent* seperti *neural networks* dan *symbolic learning*. Namun, sangat sedikit penelitian yang menggunakan *evolutionary algorithm* seperti Algoritma Genetika (GA) untuk diterapkan pada *Information Retrieval* (IR)[7].

Algoritma Genetika (GA) merupakan suatu algoritma pencarian yang berdasarkan pada mekanisme sistem natural yakni genetik dan seleksi alam. Berbeda dengan teknik pencarian konvensional, Algoritma Genetika (GA) berangkat dari himpunan solusi yang dihasilkan secara acak. Himpunan ini disebut dengan populasi. Sedangkan setiap individu dalam populasi disebut kromosom yang merupakan representasi dari solusi. Kromosom-kromosom berevolusi dalam suatu proses iterasi yang berkelanjutan yang disebut generasi. Pada setiap generasi, kromosom dievaluasi berdasarkan suatu fungsi evaluasi (fungsi *fitness*). Solusi dari generasi ini akan mengembalikan satu kromosom terbaik yang diharapkan merupakan solusi optimal.

Dalam *original matching function* yang menggunakan *vector space model*, pencarian *relevant documents* didasarkan pada tingkat kesamaan (*similarity measurement*) antara *query* inputan *user* dan *collection documents* yang ada. Dalam Algoritma Genetika (GA), setiap *query* akan direpresentasikan sebagai sebuah *chromosome*. *Chromosome* ini akan melalui proses berulang dalam *genetic operator* yaitu *selection*, *crossover*, dan *mutation* sampai didapatkan *query chromosome* terbaik untuk pengambilan dokumen yang dianggap paling *relevant*[4]. Model dengan perluasan metode ini diharapkan dapat memberikan nilai IAP yang lebih baik bila dibandingkan dengan *original matching function*, yaitu model *vector space model* tanpa penerapan optimasi *query*.

Algoritma Genetika (GA) diterapkan dalam IR untuk membentuk formula *query* baru melalui *pseudo relevance feedback* yang akan digunakan untuk pengambilan dokumen yang dianggap paling *relevant*[4].

## 1.2 Perumusan Masalah

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

1. Bagaimana proses pencarian calon *additional keywords* dari *relevant documents* yang ada menggunakan *pseudo relevance feedback*.
2. Bagaimana proses pencarian *query chromosome* paling optimal dengan menggunakan *proposed method*.
3. Bagaimana nilai perbandingan *Precision*, *Recall*, dan IAP antara metode pada Tugas Akhir ini dengan *original matching function*-nya.

Batasan masalah agar tidak meluasnya materi pembahasan dalam tugas akhir ini ialah:

1. Pengujian performansi akan dilakukan dengan membandingkan nilai *Precision*, *Recall*, dan IAP yang didapat melalui metode ini dengan *original matching function*-nya yang menggunakan *vector space model* tanpa optimasi *query*.
2. Simulasi yang dibuat berbasis web menggunakan PHP dan basisdata MySQL.
3. Koleksi dokumen dan *query* yang digunakan menggunakan bahasa inggris.
4. Analisis data dokumen dilakukan terhadap standar koleksi dokumen untuk *information retrieval* yang didapat dari: <ftp://ftp.cs.cornel.edu/pub/smart/med/> yang bertipe file “.txt” dan meliputi *Med collection* dimana sudah terdapat kumpulan *query* beserta *relevance judgment* untuk setiap *query* tersebut.
5. *Query* pengujian (*query uji*) untuk pengukuran performansi sistem sudah ditentukan sebelumnya.
6. Banyak dokumen dan calon *additional keywords* yang akan digunakan hasil dari *pseudo relevance feedback* hanya diambil top 4 dokumen teratas (dengan acuan nilai *similarity* terbesar) dan top 4 kata sebagai calon *additional keywords* (dengan acuan bobot *term* terbesar).
7. Aplikasi hanya melakukan *word indexing* dan tidak melakukan *phrase indexing*.
8. Pada Algoritma Genetika, besar nilai peluang rekombinasi (*Pc*) yang digunakan ialah 0.9 dan peluang mutasi (*Pm*) ialah 0.06.
9. Pada pengujian *query expansion*, banyaknya pengujian *query expansion* untuk setiap *query uji* ialah sebanyak 5 kali.

## 1.3 Tujuan

Secara umum tujuan penulisan yang ingin dicapai dalam Tugas Akhir ini ialah:

1. Merancang dan membangun suatu *Information Retrieval System* berupa *search engine* yang menggunakan Metode Algoritma Genetika (GA) untuk mendapatkan formula *query* baru yang paling optimal.

2. Menganalisa nilai *Precision*, *Recall*, dan IAP yang didapat dan membuktikan apakah metode ini mampu memberikan hasil yang memuaskan bila dibandingkan dengan metode awalnya tanpa optimasi *query*.

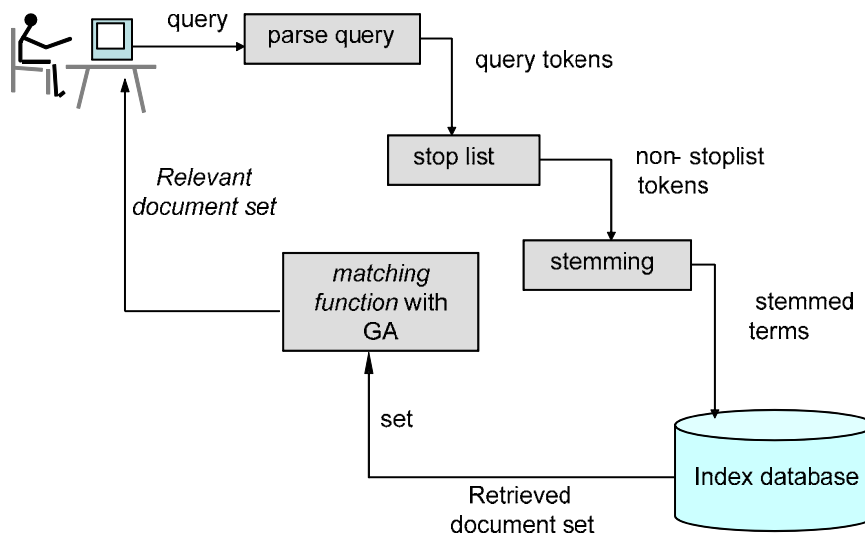
#### 1.4 Metodologi Penyelesaian Masalah

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

1. Studi Literatur  
Studi literature dari beberapa buku, jurnal, artikel yang membahas tentang *Information Retrieval* (IR), Algoritma Genetika (GA), dan penerapan GA pada IR.
2. Analisis dan Desain  
Tahap ini meliputi analisis kebutuhan serta penyelesaian masalah untuk merancang perangkat lunak *Information Retrieval System* dengan metode Algoritma Genetika (GA) untuk optimasi *query*-nya. Langkah – langkah proses secara umum dapat dituliskan sebagai:
  - i. User menginputkan *query* awal ke dalam system.
  - ii. System memberikan *retrieved document* dari database.
  - iii. Encode dokumen menjadi *chromosome* (*initial population*).
  - iv. Proses perulangan iterasi generasi yang terdiri dari *selection*, *crossover*, dan *mutation* hingga tercapai *optimize query chromosome*.
  - v. Decode *optimize query chromosome* menjadi formula *query* baru dan gunakan untuk mencari *relevant documents* dari database.

Desain perangkat lunak yang akan dibangun berdasarkan proses berikut:

- a. Proses pencarian *relevant documents* dengan menerapkan Algoritma Genetika (GA) dalam *matching function* IR:



Gambar 1-1. *Search Subsystem* IR dengan GA.

3. Implementasi Sistem  
Tahap ini meliputi pembangunan perangkat lunak yang telah dirancang pada tahap sebelumnya. Pembangunan perangkat lunak berbasis web dengan menggunakan PHP dan *database MySQL*.
4. Analisis dan Pengujian  
Pada tahap ini akan dilakukan pengujian terhadap perangkat lunak yang telah dibangun serta melakukan analisis terhadap hasil *relevant* dokumen yang didapat. Analisa performansi dari *search engine* ini setelah diterapkannya Algoritma Genetika (GA) dalam *matching function* IR akan dinilai dari nilai IAP yang dihasilkan dari metode ini dan kemudian akan dibandingkan dengan nilai IAP dari *original matching function*-nya.
5. Penyusunan 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.