

BAB I

PENDAHULUAN

1.1. Latar Belakang

World Wide Web (WWW) memungkinkan orang untuk berbagi informasi dari *repository database* secara global. Terdapat jutaan informasi baru yang muncul dalam setiap waktunya. Oleh karena itu, kita membutuhkan alat yang secara umum dikenal dengan sebutan *search engine* (mesin pencari). Saat ini mesin pencari menjadi alat yang paling berguna untuk memperoleh informasi dari Internet. Namun, hasil pencarian dari mesin pencari seringkali tidak memuaskan. Hal ini dikarenakan WWW merupakan suatu basisdata global yang tidak memiliki struktur semantik, dan hal itu membuat mesin sulit untuk memahami informasi yang diberikan oleh pengguna[1]. Oleh karena itu, tidak jarang jika mesin pencari memberikan umpan balik atau jawaban berupa banyak *link* dari halaman Web yang tidak ada hubungannya dengan kebutuhan pengguna[2].

Mesin pencari menggunakan berbagai pendekatan untuk menutupi setiap kelemahan yang ada seperti *pattern matching*, *documents similarity*, dan berbagai teknik *machine learning* lainnya (*Bayesian*, *Decision Tree*, *Resource Limited Immune classifier System*, *k-Nearest Neighbour*, dll) [3]. Sekalipun metode-metode tersebut sudah memberikan hasil yang cukup baik tetapi tetap saja belum cukup optimal. Karena inti permasalahannya adalah, mesin pencari klasik menggunakan dokumen tidak terstruktur[4] yaitu *hypertext* [5] atau HTML pada konten-konten web pada umumnya sebagai dokumen acuan untuk pencarian[6], yang masih berbasiskan pada bahasa alami. Sehingga hal tersebut menyebabkan hasil pencarian kurang akurat dan kurang sesuai dengan kata kunci yang telah dimasukkan.

Pada pertengahan tahun 2001 telah ditemukan sebuah teknologi yang dapat menjawab setiap permasalahan diatas yaitu web semantik[6]. Mesin pencari yang menggunakan teknologi web semantik dengan kemampuan iteroperabilitas dan relasi antar *web resource*-nya[7] menjanjikan sebuah hasil pencarian yang lebih akurat. Selain itu, yang menjadi kelebihan dari web semantik adalah memfokuskan objek pencarian pada sebuah dokumen standar bukan hanya pada sintaksisnya

saja tetapi juga pada isi webnya yaitu *Resource Description Framework* (RDF)[8]. RDF adalah bahasa untuk mewakili informasi tentang *resource* di WWW. Informasi dalam RDF digambarkan kedalam sebuah *triples* (subjek/*resource*, predikat/*property*, objek/*value*) [9].

Domain web pada umumnya merupakan domain dengan dokumen web belum terstruktur, masih menggunakan dokumen *hypertext* dan belum terstandarisasikan dalam bentuk RDF. Sehingga dalam desain dan implementasi mesin pencari berbasis web semantik ini terlebih dahulu perlu dibuat sebuah dokumen RDF dari setiap konten yang termuat WWW.

1.2. Perumusan Masalah

Berdasarkan uraian latar belakang, terdapat beberapa permasalahan yang dapat dirumuskan sebagai berikut :

1. Bagaimana konversi dokumen web HTML kedalam bentuk RDF?
2. Bagaimana mengoptimisasi pencarian semantic web menggunakan Jena dan SPARQL?
3. Bagaimana membuat sebuah *prototype* mesin pencari yang mampu dipakai sebagai alat uji coba keberhasilan sistem?

1.3. Tujuan dan Manfaat

Penelitian ini diharapkan mampu :

1. Menghasilkan dokumen RDF dari konversi database hasil *crawling* HTML .
2. Menghasilkan optimalisasi pencarian semantic web menggunakan Jena dan SPARQL.
3. Menghasilkan *protorype* mesin pencari berbasis web semantik.

1.4. Batasan Masalah

Agar tercapai tujuan penelitian dengan baik, maka perlu adanya pembatasan ruang lingkup penelitian, yaitu :

1. Penelitian ini menggunakan studi kasus domain detikinet.com.
2. Penelitian ini tidak menangani pembuatan *web crawler* domain detikinet.com.

3. Implementasi teknologi web semantic menggunakan rekomendasi standar *World Wide Web Consortium (W3C)* yaitu RDF, Jena dan SPARQL.
4. Proses pencarian menggunakan modul ARQ pada Jena dan SPARQL yang ditulis dengan bahasa pemrograman Java.

1.5. Metodologi

Secara garis besar, metodologi penelitian ini adalah:

1. Studi literatur

Literatur yang digunakan mencakup buku teks, jurnal ilmiah, artikel dan sumber lain baik dalam bentuk *hardcopy* maupun *online* yang terkait dengan mesin pencari dan web semantik.

2. Survei produk mesin pencari, web semantik, dan *website* universitas lain yang nantinya akan digunakan sebagai acuan dalam implementasi penelitian ini. Mesin pencari seperti Google, Yahoo, dan Bing.
3. Berdasarkan studi literatur dan survei yang telah dilakukan sebelumnya, maka dapat memulai menganalisis dan merancang mesin pencari berbasis web semantik.
4. Implementasi dan pengujian sistem, berdasarkan analisis dan rancangan yang telah dilakukan maka dapat dimulai *coding*, dan pengujian sistem.
5. Penarikan kesimpulan, hasilnya dianalisis untuk menarik kesimpulan mengenai fungsionalitas dan kinerja mesin pencari yang telah diimplementasikan.

1.6. Sistematika Penulisan

Sistematika penulisan yang digunakan untuk menyelesaikan tugas akhir adalah sebagai berikut:

BAB I : PENDAHULUAN

Menjelaskan mengenai latar belakang, permasalahan, batasan masalah, tujuan, metodologi penelitian, dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA

Membahas beberapa teori penunjang yang berhubungan dengan pokok pembahasan dalam tugas akhir ini yang secara garis besar berisi tentang penerapan teknologi web semantik pada mesin pencari.

BAB III : PERANCANGAN SISTEM

Membahas mengenai desain dan perancangan sistem. Desain sistem meliputi desain data, desain proses dan desain antar muka.

BAB IV : IMPLEMENTASI DAN UJI COBA

Pada bab ini akan dijelaskan mengenai implementasi pembuatan aplikasi berdasarkan hasil rancangan yang telah dibuat sebelumnya. Aplikasi yang telah selesai ini nantinya juga akan diuji dan dievaluasi untuk diperbaiki demi kelayakan sistem dan keberhasilan dari sistem ini sesuai dengan tujuannya dibangun.

BAB V : PENUTUP

Bab ini berisi tentang kesimpulan yang dapat diambil dari pelaksanaan tugas akhir ini, serta saran untuk pengembangan sistem.