BAB I

PENDAHULUAN

I.1 Latar Belakang Masalah

Perkembangan teknologi dewasa ini semakin memudahkan pengguna mendapatkan informasi yang diinginkan. Salah satu sumber yang mudah digunakan untuk mendapatkan informasi tersebut adalah *search engine* (mesin pencari) dan *social network* (jejaring sosial). Mesin pencari adalah salah satu tempat di mana pengguna dapat mencari informasi apapun dengan mudah. Pengguna hanya tinggal mengetikan kata kunci yang ingin dicari dan beberapa saat kemudian hasil pencarian berdasarkan kata kunci tersebut dapat terlihat. Jejaring sosial adalah tempat dimana seseorang dapat saling berinteraksi dengan orang lain secara tidak langsung. Dengan adanya mesin pencari dan jejaring sosial, maka pengguna dapat meng-*update* informasi kapan pun dan dimanapun. Mesin pencari dan jejaring sosial sendiri bukanlah hal yang asing di Indonesia, karena berdasarkan hasil survei yang dilakukan oleh internetworldstats.com [7], sebelum 31 Desember 2011, pengguna internet di Indonesia sudah mencapai angka 55.000.000 orang. Sedangkan semiostat.com [8], merilis angka akun pengguna Twitter di Indonesia yang dibuat sebelum tanggal 1 Januari 2012 mencapai angka 19.500.000 akun. Angka ini menempatkan Indonesia sebagai negara ke-5 pengguna Twitter terbanyak di dunia.

Salah satu cara yang dapat dilakukan pengguna mesin pencari dan jejaring sosial tersebut untuk mendapatkan informasi adalah melakukan pencarian

terhadap topik yang ingin diketahui. Pencarian tersebut biasanya akan menghasilkan hasil penelusuran yang tidak sedikit, mulai dari ratusan sampai jutaan hasil akan ditampilkan. Namun, hasil teratas pencarian tersebut tidak selamanya relevan dengan informasi yang benar-benar diinginkan. Tidak jarang, informasi yang benar-benar sesuai justru tersimpan di bagian bawah dari halaman hasil pencarian, yang kemungkinan besar tidak akan ditelusuri pengguna. Hal ini berlaku untuk mesin pencari sekelas Google sekalipun yang pada tahun 2010 menempati urutan pertama sebagai "*top most search engine for 2010*" [9], seperti dikemukakan dalam [10]. Walaupun Google sendiri telah memiliki solusi terhadap masalah ini, yaitu dengan adanya algoritma pencarian baru yang bernama Google Panda, tetap saja algoritma tidak memperbaiki hasil pencarian seratus persen akurat [12].

Salah satu solusi untuk permasalahan di atas adalah dengan mengorganisasikan hasil pencarian yang didapat ke dalam kelompok-kelompok bahasan berdasarkan topik yang dimiliki. Pengguna dapat melihat dan menelusuri ringkasan dari hasil pencarian tersebut secara lebih cepat. Solusi ini diharapkan dapat membuat waktu yang digunakan pengguna lebih efektif dan efisien saat melakukan penelusuran dari hasil pencarian terhadap suatu informasi tertentu. Penerapan solusi ini dapat dilakukan dengan metode *clustering*.

Clustering adalah pengelompokan objek yang berdasarkan kemiripan yang dimiliki ke dalam suatu grup. *Clustering* melakukan pengelompokan dokumen tanpa ditentukan terlebih dahulu kelompok-kelompok atau kelas-kelas penampungnya. *Clustering* sangat berguna dalam pengelompokan dokumen,

karena dapat mengekstrak topik dari dokumen secara otomatis dan tanpa perlu ditentukan terlebih dahulu *cluster* apa saja yang harus ada untuk menampungnya. *Clustering* sangat berguna untuk pengelompokan dokumen hasil pencarian *online* karena hasil pencarian yang ditampilkan bisa sangat bermacam-macam walaupun dalam satu kata kunci. Diharapkan pengguna dapat melakukan seleksi dan menemukan informasi lebih cepat dari hasil penelusuran yang telah dikelompokan.

Suffix Tree Clustering (STC) adalah metode yang lebih baru untuk *clustering*. STC menggunakan frasa sebagai pembangun *cluster*, dan oleh karena itu dianggap lebih sederhana jika dibandingkan dengan algoritma lain. Misalnya jika dibandingkan dengan algoritma *Agglomerative Hierarchical Clustering* (AHC) yang cukup sering digunakan, performansi AHC akan lebih lambat jika diaplikasikan ke dalam set dokumen yang banyak dan AHC membutuhkan konstanta dalam penentuan berapa *cluster* yang harus terbentuk. Salah satu keunggulan STC lain dibanding AHC adalah dapat membantu mengelompokan topik-topik yang *overlap*. Topik yang *overlap* adalah topik yang dapat berada di lebih dari satu *cluster*. Salah satu keunggulan STC yang lain adalah STC memiliki keunggulan waktu eksekusi yang cepat, yang biasanya didapat dengan cara menentukan jumlah *cluster* yang dibutuhkan terlebih dahulu. Permasalahan-permasalahan di atas menjadi dasar penulis tertarik untuk melakukan penelitian terhadap implementasi STC untuk pengelompokan dokumen hasil pencarian secara *online* pada mesin pencari dan jejaring sosial. Studi kasus dilakukan di mesin pencari Google dan jejaring sosial Twitter. Google dipilih karena Google

merupakan mesin pencari terpopuler [9], sedangkan Twitter dipilih karena dinilai lebih informatif berdasarkan batasan karakter yang dapat ditulis dalam *tweet*.

I.2 Perumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah yang dikaji dalam penelitian ini adalah bagaimana menerapkan metode *Suffix Tree Clustering* untuk pengelompokan dokumen hasil pencarian secara *online* pada mesin pencari dan jejaring sosial.

I.3 Maksud dan Tujuan

Berdasarkan permasalahan yang diteliti, maka maksud dari penulisan penelitian ini adalah untuk mengimplementasikan metode *Suffix Tree Clustering* untuk pengelompokan dokumen hasil pencarian secara *online*.

Sedangkan tujuan yang akan dicapai dalam penelitian ini adalah :

1. Untuk menganalisis metode *Suffix Tree Clustering* pada pengelompokan dokumen secara umum.
2. Untuk mengimplementasikan metode *Suffix Tree Clustering* pada pengelompokan dokumen hasil pencarian *online*.
3. Untuk mendapatkan persamaan dan perbedaan pola dokumen, sehingga dapat diperoleh ringkasan yang berguna untuk pengelompokan.

I.4 Batasan Masalah

Untuk memberikan suatu pandangan dan arah yang jelas dalam penelitian ini, maka penulis membatasi masalah sebagai berikut :

1. Studi kasus dilakukan di mesin pencari Google dan jejaring sosial Twitter.
2. Sumber data berasal dari hasil pencarian dokumen berdasarkan suatu kata kunci.
3. Metode yang digunakan adalah metode *Suffix Tree Clustering*.
4. *Clustering* data dilakukan pada halaman hasil yang telah ditelusuri oleh mesin pencari Google dan Twitter.. Hasil *clustering* akan berada di samping isi halaman hasil pencarian.
5. Semua tanda kutip (“) dan petik (‘) yang terdapat dalam hasil pencarian akan dirubah menjadi garis bawah (_) setelah proses *clustering* dilakukan.
6. Metode analisis dan pembangunan perangkat lunak yang digunakan adalah metode analisis terstruktur.

I.5 Metodologi Penelitian

Metodologi yang digunakan dalam penulisan penelitian ini adalah metode analisis deskriptif, dengan metode-metode sebagai berikut :

1. Metode pengumpulan data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah sebagai berikut :

a. Studi Literatur.

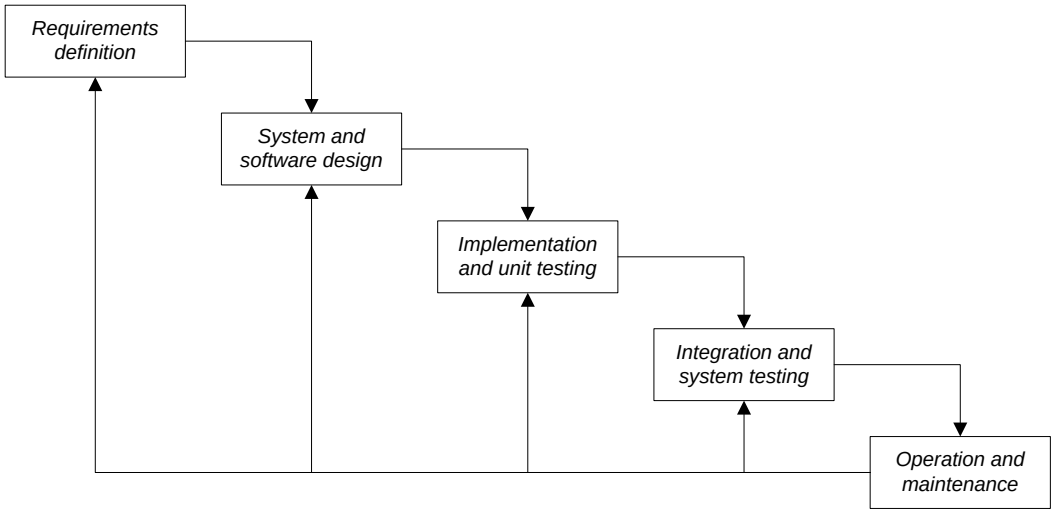
Suatu metode pengumpulan data dengan cara mengumpulkan literatur, jurnal, *paper*, dan bacaan-bacaan yang ada kaitannya dengan penelitian.

b. Observasi.

Suatu teknik pengumpulan data dengan melakukan penelitian dan peninjauan langsung terhadap permasalahan yang diambil.

2. Metode pembangunan perangkat lunak.

Teknik analisis data dalam pembuatan perangkat lunak menggunakan paradigma perangkat lunak secara *waterfall* [13] (Gambar I.1), yang meliputi beberapa proses diantaranya :



Gambar I-1 Metode *Waterfall*

a. Requirements Definitions

Seluruh kebutuhan *software* harus sudah didapatkan dalam fase ini, termasuk di dalamnya kegunaan *software* yang diharapkan pengguna dan batasan *software*. Informasi ini biasanya dapat diperoleh melalui wawancara, survey atau diskusi.

Informasi tersebut dianalisis untuk mendapatkan dokumentasi kebutuhan pengguna untuk digunakan pada tahap selanjutnya.

b. System and Software Design

Tahap ini dilakukan sebelum melakukan *coding*. Tahap ini bertujuan untuk memberikan gambaran apa yang seharusnya dikerjakan dan bagaimana tampilannya. Tahap ini membantu dalam menspesifikasi kebutuhan *hardware* dan sistem serta mendefinisikan arsitektur sistem secara keseluruhan.

c. Implementation and Unit Testing

Dalam tahap ini dilakukan pemrograman. Pembuatan *software* dipecah menjadi modul-modul kecil yang nantinya akan digabungkan dalam tahap berikutnya. Selain itu dalam tahap ini juga dilakukan pemeriksaan terhadap modul yang dibuat, apakah sudah memenuhi fungsi atau belum.

d. Integration and System Testing

Di tahap ini dilakukan penggabungan modul-modul yang sudah dibuat dan dilakukan pengujian. Ini dilakukan untuk mengetahui apakah *software* yang dibuat telah sesuai dengan desainnya dan masih terdapat kesalahan atau tidak.

e. Operation and Maintenance

Ini merupakan tahap terakhir dalam model *waterfall*. *Software* yang sudah jadi dijalankan serta dilakukan pemeliharaan. Pemeliharaan termasuk dalam memperbaiki kesalahan yang tidak ditemukan pada langkah sebelumnya. Perbaikan implementasi unit sistem dan peningkatan jasa sistem sebagai kebutuhan baru juga dilakukan pada tahap ini.

I.6 Sistematika Penulisan

Sistematika penulisan proposal penelitian ini disusun untuk memberikan gambaran umum untuk memudahkan dalam menganalisa dan memahami tentang penelitian yang dilakukan. Sistematika penulisan penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini menguraikan tentang latar belakang permasalahan, rumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian yang digunakan, serta sistematika penulisan dalam penelitian ini.

BAB II TINJAUAN PUSTAKA

Bab ini membahas berbagai konsep dasar dan teori-teori yang berkaitan dengan topik penelitian yang dilakukan dan hal-hal yang berguna dalam proses analisis permasalahan. Teori-teori yang dibahas mengenai perangkat lunak, *data flow diagram*, *entity relationship diagram*, *data mining*, *clustering*, *text mining*, *suffix tree clustering*, dan aplikasi pembangun perangkat lunak.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Bab ini membahas mengenai dua bagian utama, yaitu analisis dan perancangan sistem yang akan dibangun. Pada tahap analisis terdapat lima bagian yang akan dijelaskan, yaitu mengenai masalahnya, sistem yang sedang berjalan, spesifikasi kebutuhan perangkat lunak, analisis kebutuhan non-fungsional, serta kebutuhan fungsional. Sedangkan pada tahap perancangan terdapat empat bagian yang akan dijelaskan, yaitu penjelasan mengenai perancangan antarmuka, perancangan struktur menu, jaringan semantik, dan perancangan prosedural.

BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini membahas dua bagian utama, yaitu implementasi dan pengujian. Pada bagian implementasi terdapat empat bagian yang akan dijelaskan, yaitu penjelasan mengenai perangkat keras, perangkat lunak, basis data serta antarmuka. Sedangkan pada bagian pengujian akan dijelaskan mengenai pengujian yang berupa pengujian *alpha* dengan menggunakan metode *blackbox* serta pengujian *beta* dengan menggunakan kuesioner.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan berdasarkan hasil penelitian yang telah dilakukan dan masukan-masukan yang dapat digunakan untuk pengembangan penelitian selanjutnya.