BAB I
PENDAHULUAN

1.1 Latar Belakang

Pada saat ini teknologi sudah banyak berkembang terutama dalam hal penyimpanan data. Data yang ditampung terlalu banyak dan dalam jumlah yang besar menimbulkan masalah baru terutama dalam hal pencarian atau dalam hal pengaksesan data. kadang kala pemrosesan data menggunakan cara manual sudah tidak digunakan karena tidak efektif dan efisien contohnya dalam hal pencarian sebuah kata dalam document yang besar atau dalam hal penyaringan kata. Beberapa masalah yang dihadapi diantaranya penyaringan kata dalam setiap dokumen yang terdiri dari berbagai kata dasar yang sulit diseleksi, apalagi jiga kata dasar tersebut disertai dengan imbuhan atau serapan dari bahasa asing, atau penyaringan kata setiap dokumen yang memiliki stop word berupa kata yang tidak memiliki makna dan tidak diperlukan serta dalam hal perhitungan seberapa sering kata kunci muncul dalam setiap dokumen.Teknik-teknik Text-Mining digunakan untuk menemukan informasi dan knowledge yang bisa diambil dari sekumpulan data yang sangat besar tersebut.

Text mining adalah suatu proses analisis dalam sekumpulan data dalam jumlah besar berdasarkan informasi berkualitas tinggi dari sebuah teks dan tujuannya adalah mencari kata-kata yang dapat mewakili dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen. Tahapanyang terjadi dalam melakukan analisis text mining adalah dengan pembacaan data kemudian ektraksifitur yang bertujuan untuk mengekstrak kata-kata kunci dari dokumen kemudian dilakukannya proses *tokenization* yang berfungsi untuk pemecah text menjadi kata tunggal kemudian proses selanjutnya adalah filtering yang bertujuan untuk mengambil kata-kata penting dari hasil tokenizing kemudian dilakukan proses stemming yang berfungsi untuk mencari root kata dari setiap kata hasil filtering kemudian tagging yang bertujuan untuk mencari bentuk awaldari tiap kata lampau dari hasil stemming kemudian proses selanjutnya stopword removing yang bertujuan menghilangkan kata tidak penting dalam text menjadi kata dasar.

1.2 Rumusan Masalah

Bagaimana membangun sebuah aplikasi Text Mining di Lembaga Ilmu Pengetahuan Indonesia (LIPI) berbasis desktop.

1.3 Maksud Dan Tujuan

Maksud dari penelitian ini adalah untuk membuat aplikasi Text Mining di Lembaga Ilm Pengetahuan Indonesia berbasis desktop. Sedangkan tujuan yang akan dicapai dalam penelitian ini adalah:

1. Membuat aplikasi text mining untuk mempermudah proses pemilihan kata-kata dari setiap dokumen.
2. mengimplementasikan metode penghapusan kata dasar dari sebuah document
3. menghilangkan stopword ataua kata yang tidak penting yang sering ditemukan dalam sebuah dokument.
4. Menghitung frequensi kemunculan setiap kata dalam document yang besar

1.4 Batasan Masalah

Dalam penelitian ini, penulis membatasi masalah sebagai berikut :

1. Data yang diolah pada aplikasi ini adalah :
   a. Data dokumen yang berextensi dot(.) txt
   b. Data inputan manual oleh user
2. Analisis dan pemodelan yang digunakan dalam pembangunan aplikasi adalah UML.

1.5 Metode Penelitian

Metode penelitian merupakan tahapan-tahapan yang dilalui oleh peneliti mulai dari perumusan masalah sampai kesimpulan, yang membentuk alur yang sistematis. Metode penelitian ini digunakan sebagai pedoman peneliti dlam melaksanakan penelitian ini agar hasil yang dicapai ini tidak menyimpang dari tujuan yang telah ditentukan

sebelumnya. Metode penelitian penelitian yang dilakukan terdiri dari metode pengumpulan data dan metode pembangunan perangkat lunak.

1.5.1 metode pengumpulan data

metode pengumpulan data yang digunakan dalam penelitian ini adalah :

1. studi literature

   studi ini dilakukan dengan cara mempelajari, meneliti dan menelaah berbagai literature dari perpustakaan yang bersumber buku-buku,jurnal ilmiah,situs-situs internet yang ada kaitannya dengan judul penelitian.

1.5.2 metode pembangunan perangkat lunak

metode yang digunakan dalam pembuatan perangkat lunak ini menggunakan model *waterfall*. Berikut ini adalah gambaran dan fase-fase dari model model *waterfall* menurut Roger S.Presman.

penjelasan model waterfall adalah sebagai berikut :

a. communication

   langkah ini merupakan analisis terhadap kebutuhan *software*, dan tahap untuk mengadakan pengumpulan data dengan melakukan pertemuan dengan *costumer*, maupun mengumpulkan data-data tambahan baik yang ada dijurnal maupun dari internet.

b. Planning

   Proses *planning* merupakan lanjutan dari proses *communication* (*analysis requirement*). Tahapan ini akan menghasilkan dokumen *user requirement* atau bias dikatakan sebagai data yang berhubungan dengan keinginan *user* dalam pembuatab *software*, termasuk rencana yang akan dilakukan.

c. Modeling

   Proses *modeling* ini akan menerjemahkan syarat kebutuhan ke sebuah perancangan *software* yang dapat diperkirakan sebelum dibuat *coding*, proses ini berfokus pada rancangan struktur data, arsitektur *software*, representasi *interface*, dan detail (algoritma)

procedural. Tahapan ini akan menghasilkan dokumen yang disebut *software requirement*.

d. Construction

Construction merupakan proses membuat kode. *Coding* atau pengkodean merupakan penerjemah desain dalam bahasa yang bias dikenali oleh komputer. *Programmer* akan menerjemahkan transaksi yang diminta oleh *user*. Tahapan inilah yang merupakan tahapan secara nyata dalam mengerjakan suatu *software*, artinya penggunaan komputer akan dimaksimalkan dalam tahapan ini. Setelah pengkodean selesai maka akan dilakukan testing terhadap system yang telah dibuat. Tujuan testing adalah menemukan kesalahan-kesalahan terhadap system tersebut untuk kemudian bias diperbaiki.

e. Deployment

Tahapan ini merupakan *final* dalam pembuatan sebuah *software* atau system. Setelah melakukan analisis, desain dan pengkodean maka system yang sudah jadi akan digunakan oleh *user*. Kemudian *software* yang telah dibuat harus dilakukan pemeliharaan secara berskala.

1.6 Sistematika Penulisan

Sistematika dari penulisan laporan kerja praktek ini dibagi menjadi empat bab pembahasan yaitu sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini akan dijelaskan mengenai latar belakang permasalahan, rumusan masalah, maksud dan tujuan, batasan masalah, metode penelitian dan sistematika penulisan.

BAB II Gambaran umum instansi

Pada bab ini akan dijelaskan mengenai sejarah instansi,ruang lingkup kerja,landasan teori dan analisis perancangan perangkat lunak.

BAB III PERANCANGAN

Pada bab ini akan membahas tentang perancangan sistem aplikasi *text mining*.

BAB IV SIMPULAN DAN SARAN

ada bab ini berisi pengujian dan pembahasan tentang implementasi dari penerapan teknik *data mining* serta validasi aplikasi yang penulis buat.

Bab ini berisi tentang kesimpulan dari pembuatan aplikasi *text mining* menggunakan algoritma poreter dan berisi saran-saran yang ditujukan kepada semua pihak yang bersangkutan.