BAB 2

TINJAUAN PUSTAKA

2.1 Profil Balai Bahasa Bandung

Berikut ini adalah profil dari Balai Bahasa Bandung yang merupakan tempat penelitian untuk membangun aplikasi *Stemming Engine*.

2.1.1 Sejarah Balai Bahasa Bandung

Pasal 36 UUD 1945 menyebutkan bahwa bahasa resmi negara adalah bahasa Indonesia. Berlandaskan pada pasal itu Pemerintah Indonesia berusaha melestarikan, membina, dan mengembangkan bahasa Indonesia dan bahasa daerah. Sebagai tindak lanjutnya, pemerintah membentuk lembaga yang bernama Pusat Bahasa (sebelumnya pada tahun 1966 bernama Lembaga Bahasa dan Budaya, kemudian diubah menjadi Lembaga Bahasa Nasional pada tahun 1969. Sejak April 1975 Lembaga Bahasa Nasional diubah menjadi Pusat Pembinaan dan Pengembangan Bahasa, lalu nama tersebut diubah lagi menjadi Pusat Bahasa pada tahun 1999). Lembaga ini berada di bawah Departemen Pendidikan Nasional.

Pusat Bahasa mempunyai tugas utama mengembangkan dan membina bahasa dan sastra Indonesia dan daerah. Mengingat wilayah yang harus dijangkau meliputi seluruh Indonesia, Pusat Bahasa mendirikan balai-balai bahasa di seluruh Indonesia. Balai bahasa yang berada di Provinsi Jawa Barat adalah Balai Bahasa Bandung (yang nantinya akan bernama Balai Bahasa Provinsi Jawa Barat). Balai Bahasa Bandung ini resmi berdiri pada bulan September 1999.

2.1.2 Visi Dan Misi Balai Bahasa Bandung

Berikut ini adalah visi dan misi Balai Bahasa Bandung:

Visi

Visi Balai Bahasa Bandung adalah terwujudnya lembaga penelitian yang unggul dan pusat informasi serta pelayanan di bidang kebahasaan dan kesastraan (Indonesia dan daerah) di Jawa Barat.

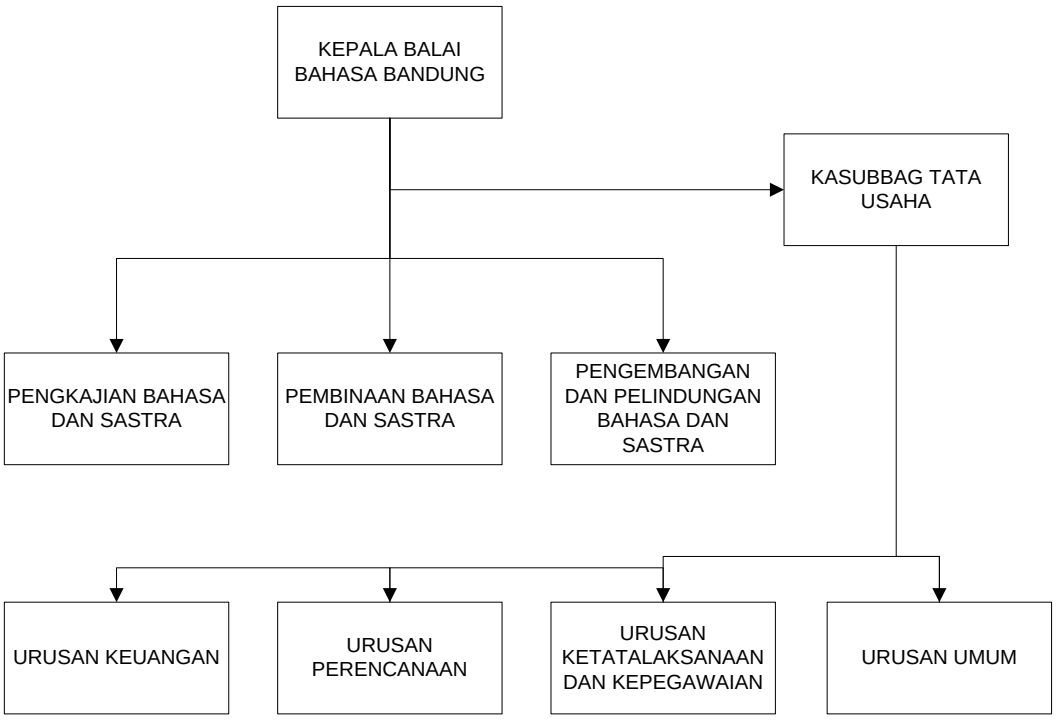
Misi

Misi Balai Bahasa Bandung adalah:

1. meningkatkan mutu bahasa dan sastra;
2. meningkatkan sikap positif masyarakat terhadap bahasa dan sastra;
3. mengembangkan bahan informasi kebahasaan dan kesastraan;
4. mengembangkan tenaga kebahasaan dan kesastraan;
5. meningkatkan kerja sama.

2.1.3 Struktur Organisasi Balai Bahasa Bandung

Struktur organisasi Balai Bahasa Bandung tidaklah rumit dan kompleks hanya terdiri dari bagian yang penting saja misalnya Kepala, Sub Bagian, serta kelompok Jabatan Fungsional. Untuk lebih jelasnya dapat dilihat dari gambar berikut.



Gambar 2.1 Struktur Organisasi Balai Bahasa Bandung

2.1.4 *Job Description*

Berikut ini adalah *job description* yang terdapat pada Balai Bahasa Bandung di bagian kelompok jabatan fungsional :

1. Menyusun program rencana penelitian, pembinaan dan pengembangan bahasa dan sastra di daerah.
2. Mengkaji, menganalisis, menyusun bahan kebijakan teknis pembinaan bahasa dan sastra di daerah.
3. Mengkaji, menganalisis, dan meyusun konsep, pedoman, standar dan norma pembinaan bahasa dan sastra di daerah.
4. Melakukan penyusunan penelitian, pembinaan dan pengembangan bahasa dan sastra di daerah.
5. Melakukan inventarisasi masalah bahasa dan sastra di daerah.

6. Melakukan penelitian bahasa dan sastra di daerah.
7. Melakukan evaluasi hasil penelitian, pembinaan, dan pengembangan bahasa dan sastra di daerah.
8. Menyajikan data dan informasi jaringan kebahasan dan kesastraan di daerah.

2.1.5 Deskripsi Tugas dan Fungsi Balai Bahasa Bandung

Tugas pokok Balai Bahasa Bandung adalah melaksanakan penelitian, pengembangan, pembinaan bahasa dan sastra di daerah. Sementara itu, fungsi Balai Bahasa Bandung adalah:

1. Melaksanakan kebijakan teknis Pusat Bahasa di bidang pembinaan dan pengembangan bahasa dan sastra Indonesia.
2. Merumuskan dan melaksanakan kebijakan teknis di bidang pembinaan dan pengembangan bahasa dan sastra Indonesia di daerah.
3. Bekerja sama dengan pemerintah provinsi, kota, dan kabupaten.

2.2 Landasan Teori

Dalam pembangunan *Stemming Engine* terdapat beberapa landasan teori yang mendukung, diantaranya akan dijelaskan dibawah ini.

2.2.1 Text Mining

Text mining merupakan salah satu aplikasi dari data mining. *Text mining* juga sering disebut sebagai *Text Data Mining* (TDM) dan *knowladge Discovery in Textual Databases* (KDT). *Text mining* merupakan proses mengesktrak petterns dan *knowladge* yang bersifat menarik dan *nontrivial* (penting) dari dokumen-

dokumen teks. Pada intinya proses kerja text mining sama dengan proses kerja data mining pada umumnya hanya saja data yang di mining merupakan *text databases*.

Data teks akan diproses menjadi data numerik agar dapat dilakukan proses lebih lanjut. Sehingga dalam *text mining* ada istilah preprocessing data, yaitu proses pendahulu yang diterapkan terhadap data teks yang bertujuan untuk menghasilkan data numerik.

Pada proses preprosesing merupakan tahap dimana deskripsi di tangani untuk dapat siap diproses memasuki tahap *text mining*. Tahap-tahap tersebut adalah:

1. *Tokenizing*
2. *Filtering*
3. *Stemming*
4. *Tagging*
5. *Anayizing*

2.2.1.1 *Tokenizing*

Tokenizing adalah sebuah proses yang dilakukan seseorang untuk menjadikan sebuah kalimat menjadi lebih bermakna atau berada dengan cara memecah kalimat tersebut menjadi kata-kata atau *frase-frase ("Parsing")*. *Parsing* di dalam pembuatan aplikasi *text mining* ini merupakan proses penguraian deskripsi yang semula berupa kalimat-kalimat berisi kata-kata dan tanda pemisah antara kata seperti titik(.), koma(,), spasi dan tanda pemisah lain menjadi kata-kata saja baik itu berupa kata-kata penting maupun kata-kata tak penting. Secara sederhana proses

parsing ini terlihat sebagai proses pengambilan kata jika ketemu tanda spasi namun pada kenyataannya tidak sesederhana itu.

2.2.1.2 *Filtering*

Proses *filtering* merupakan proses mengambil kata-kata penting yang terdapat dari hasil *tokenizing*. Untuk melakukan *filtering* bisa menggunkan algoritma *stop list* atau *word list.* Pada aplikasi yang akan dibangung menggunakan algoritma *stop list* yaitu membuang kata yang kurang penting yang terdapat dalam data hasil *tokenizing*. Data hasil *tokenizing* akan dibandingkan dengan daftar kata tidak penting. Jika terdapat dalam daftar tersebut maka kata tersebut maka akan dihapus.

2.2.1.3 *Stemming*

Stemming merupakan bagian yang tidak terpisahkan dalam *Information Retrieval* (*IR*). Tidak banyak algoritma yang dikhususkan untuk *stemming* bahasa Indonesia dengan berbagai keterbatasan didalamnya. Algoritma Porter salah satunya, Algoritma ini membutuhkan waktu yang lebih singkat dibandingkan dengan *stemming* menggunakan Algoritma Nazief & Adriani, namun proses *stemming* menggunakan Algoritma Porter memiliki prosentase keakuratan (presisi) lebih kecil dibandingkan dengan stemming menggunakan Algoritma Nazief & Adriani. Algoritma Nazief & Adriani sebagai algoritma *stemming* untuk teks berbahasa Indonesia yang memiliki kemampuan prosentase keakuratan (presisi) lebih baik dari algoritma lainnya. Algoritma ini sangat dibutuhkan dan menentukan dalam proses IR dalam dokumen Indonesia [**2**].

Stemming adalah salah satu cara yang digunakan untuk meningkatkan performa IR dengan cara mentransformasi kata-kata dalam sebuah dokumen teks ke bentuk kata dasarnya. Algoritma *stemming* untuk bahasa yang satu berbeda dengan algoritma stemming untuk bahasa lainnya. Sebagai contoh bahasa Inggris memiliki morfologi yang berbeda dengan bahasa Indonesia sehingga algoritma *stemming* untuk kedua bahasa tersebut juga berbeda. Proses *stemming* pada teks berbahasa Indonesia lebih rumit/kompleks karena terdapat variasi imbuhan yang harus dibuang untuk mendapatkan *root word*(kata dasar) dari sebuah kata [**2**].

2.2.2 Pembangun Perangkat Lunak

Berikut ini adalah pembahasan teori yang merupakan pembangun dari perangkat lunak yang akan dibangun.

2.2.2.1 HTML

Sebuah *bahasa markup* yang digunakan untuk membuat sebuah halaman web dan menampilkan berbagai informasi di dalam sebuah browser Internet. Bermula dari sebuah bahasa yang sebelumnya banyak digunakan di dunia penerbitan dan percetakan yang disebut dengan SGML *(Standard Generalized Markup Language),* HTML adalah sebuah standar yang digunakan secara luas untuk menampilkan halaman web. HTML saat ini merupakan standar Internet yang didefinisikan dan dikendalikan penggunaannya oleh World Wide Web Consortium (W3C).

HTML berupa kode-kode tag yang menginstruksikan browser untuk menghasilkan tampilan sesuai dengan yang diinginkan. Sebuah file yang

merupakan file HTML dapat dibuka dengan menggunakan browser web seperti Mozilla Firefox atau Microsoft Internet Explorer. HTML juga dapat dikenali oleh aplikasi pembuka email ataupun dari PDA dan program lain yang memiliki kemampuan browser.

HTML dokumen tersebut mirip dengan dokumen teks biasa, hanya dalam dokumen ini sebuah teks bisa memuat instruksi yang ditandai dengan kode atau lebih dikenal dengan TAG tertentu.. Tanda digunakan untuk mengaktifkan instruksi cetak tebal, diikuti oleh teks yang ingin ditebalkan, dan diakhiri dengan tanda untuk menonaktifkan cetak tebal tersebut. Secara garis besar, terdapat empat jenis elemen dari HTML:

1. *structural*. tanda yang menentukan level atau tingkatan dari sebuah teks akan memerintahkan *browser* untuk menampilkan "Golf" sebagai teks tebal besar yang menunjukkan sebagai Heading 1
2. *presentational*. tanda yang menentukan tampilan dari sebuah teks tidak peduli dengan level dari teks tersebut. Tanda presentational saat ini sudah mulai digantikan oleh CSS dan tidak direkomendasikan untuk mengatur tampilan teks,
3. *hypertext*. tanda yang menunjukkan pranala ke bagian dari dokumen tersebut atau pranala ke dokumen lain akan menampilkan sebagai sebuah hyperlink ke URL tertentu),
4. Elemen *widget* yang membuat objek-objek lain seperti tombol (<button>), list (<li>), dan garis horizontal (<hr>).

Selain markup *presentational*, markup yang lain tidak menentukan bagaimana tampilan dari sebuah teks. Namun untuk saat ini, penggunaan tag HTML untuk menentukan tampilan telah dianjurkan untuk mulai ditinggalkan dan sebagai gantinya digunakan *Cascading Style Sheets*.

2.2.2.2 PHP

PHP (Personal Home Page) merupakan script untuk pemrograman web server-side, script yang membuat dokumen HTML secara on the fly, dokumen HTML yang dihasilkan dari suatu aplikasi bukan dokumen HTML yang dibuat dengan menggunakan teks editor atau editor HTML. Dengan menggunakan PHP, maintanance suatu situs web menjadi lebih mudah. Proses update data dapat dilakukan dengan menggunakan aplikasi yang dibuat dengan menggunakan script PHP. PHP/FI merupakan nama awal dari PHP, dibuat pertama kali oleh Rasmus Lerdoff. PHP awalnya merupakan program CGI yang dikhususkan untuk menerima input melalui form yang ditampilkan dalam browser web. Software ini disebarkan dan dilisensikan sebagai perangkat lunak Open Source. Adapun keunggulan yang dimiliki oleh PHP adalah:

1. *Life Cycle* yang sangat singkat, sehingga PHP selalu up to date mengikuti perkembangan teknologi internet.
2. *Cross Platform*, yakni PHP dapat dipakai di hampir semua webserver yang ada di pasaran (terutama Apache dan Microsoft IIS) dan dijalankan pada berbagai sistem operasi (Linux, Windows, FreeBSD).

3. PHP mendukung koneksi ke banyak database baik yang gratis maupun komersil, seperti MySQL, mSQL, Oracle, Microsoft SQL Server, Interbase, dan banyak lagi.
4. PHP bersifat open source dan gratis. Kemudahan dalam mendapatkan dokumentasi. di Internet, tidak akan sulit untuk mencari baik itu referensi, kode-kode PHP yang sudah jadi dan juga mengajukan pertanyaan pada grup-grup diskusi yang di dalamnya banyak sekali para master PHP.

2.2.2.3 MySQL

MySQL adalah sebuah implementasi dari sistem manajemen basisdata relasional (RDBMS) yang didistribusikan secara gratis dibawah lisensi GPL (General Public License). Setiap pengguna dapat secara bebas menggunakan MySQL, namun dengan batasan perangkat lunak tersebut tidak boleh dijadikan produk turunan yang bersifat komersial. MySQL sebenarnya merupakan turunan salah satu konsep utama dalam basisdata yang telah ada sebelumnya; SQL(*Structured Query Language*). SQL adalah sebuah konsep pengoperasian basis data, terutama untuk pemilihan atau seleksi dan pemasukan data, yang memungkinkan pengoperasian data dikerjakan dengan mudah secara otomatis.

Kehandalan suatu sistem basisdata (DBMS) dapat diketahui dari cara kerja pengoptimasi-nya dalam melakukan proses perintah-perintah SQL yang dibuat oleh pengguna maupun program-program aplikasi yang memanfaatkannya. Sebagai peladen basis data, MySQL mendukung operasi basisdata transaksional maupun operasi basisdata non-transaksional. Pada modus operasi non-

transaksional, MySQL dapat dikatakan unggul dalam hal unjuk kerja dibandingkan perangkat lunak peladen basisdata kompetitor lainnya. Namun demikian pada modus non-transaksional tidak ada jaminan atas reliabilitas terhadap data yang tersimpan, karenanya modus non-transaksional hanya cocok untuk jenis aplikasi yang tidak membutuhkan reliabilitas data seperti aplikasi blogging berbasis web (wordpress), CMS, dan sejenisnya. Untuk kebutuhan sistem yang ditujukan untuk bisnis sangat disarankan untuk menggunakan modus basisdata transaksional, hanya saja sebagai konsekuensinya unjuk kerja MySQL pada modus transaksional tidak secepat unjuk kerja pada modus non-transaksional.

Sebagai server database dengan konsep database modern, MySQL memiliki keistimewaan. Beberapa keistimewaan dimiliki MySQL sebagai berikut:

1. ***Portability Database***. MySQL berfungsi dengan stabil tanpa kendala, berarti berlaku pada berbagai sistem operasi seperti Windows, Linux, FreeBSD, Mac OS X Server, Solaris, Amiga, HP-Unix, dan lain-lain.
2. ***Open Source***. MySQL merupakan database open source (gratis), di bawah lisensi GPL sehingga dapat memperoleh dan menggunakannya secara cuma-cuma tanpa membayar sepersen pun.
3. ***Multiuser***. MySQL merupakan database yang dapat digunakan untuk menangani beberapa user dalam waktu bersamaan tanpa mengalami masalah. Dan memungkinkan sebuah database server MySQL dapat diakses client secara bersamaan pula.

4. ***Performance Tuning*.** MySQL mempunyai kecepatan yang cukup baik dalam menangani query query sederhana, serta mampu memproses lebih banyak SQL per satuan waktu.
5. ***Column Type Database***. MySQL didukung dengan tipe data yang sangat kompleks, seperti signed/unsigned integer, float, double, char, varchar, text, blob, data, time, datetime, timestamp, year, set serta enum.
6. ***Command dan Functions***. MySQL server memiliki operator dan fungsi secara penuh yang mendukung perintah SELECT dan WHERE dalam query.
7. ***Security***. Sistem Security pada MySQL mempunyai beberapa lapisan sekuritas seperti tingkatan subnetmask, hostname, dan izin akses user dengan sistem perizinan yang mendetil serta password terenkripsi.
8. ***Scalability dan Limits***. MySQL mempunyai kemampuan menangani database dalam skala cukup besar, dengan jumlah record lebih dari 50 juta dan 60 ribu tabel serta 5 miliar baris. Selain itu dapat menampung indeks sampai 32 indeks pada tiap tabelnya.
9. ***Connectivity*.** Adanya kemampuan MySQL melakukan koneksi dengan client menggunakan protokol TCP/IP, Unix socket (Unix), atau Named Pipes (NT).
10. ***Localization*.** Adanya kemampuan dalam mendeteksi kesalahan (*error code*) pada *client* menggunakan lebih dari dua puluh bahasa.

11. ***Interface*.** MySQL memiliki interface terhadap berbagai aplikasi dan bahasa pemograman menggunakan fungsi API (Application Programming Interface).
12. ***Clients dan Tools***. Database MySQL dilengkapi berbagai tools yang dapat digunakan untuk administrasi database.
13. ***Struktur Tabel*.** MySQL memiliki struktur tabel cukup baik serta cukup fleksibel, misalnya ketika menangani alter table.

2.2.2.4 Dreamweaver CS5

Adobe Dreamweaver merupakan program penyunting halaman web keluaran adobe system yang dulu dikenal sebagai micromedia dreamweaver keluaran micromedia. Program ini banyak digunakan oleh pengembang web karena fitur-fiturnya yang menarik dan kemudahan penggunanya.Versi terakhir Dreamweaver keluaran Adobe Systems adalah Adobe Dreamweaver CS5 yang ada dalam Adobe Creative Suite 5.

Fitur baru Dreamweaver CS5 yang makin handal untuk versi terbaru ini di antaranya adalah Integrated CMS Support, CSS Inspection, PHP custom class code hinting, dan site-specific code hinting. Semua fitur baru tersebut semakin memantapkan pengguna Adobe Dreamweaver CS5 untuk semakin mengeksplorasi dan mengeksploitasi ide kreasi pengolahan website.

Beberapa kelebihan adobe Dreamweaver CS5 :

1. Adobe Dreamweaver CS5 memiliki handling error

2. Kemampuannya membuat halaman web yang terlihat konsisten. Adobe Dreamweaver sudah terinstall beberapa template yang elegan dan menarik. Tentunya ini memudahkan anda yang ingin belajar membuat sebuah web namun anda belum mampu membuat design web sendiri. Jika anda ingin membuat lebih dari 10 halaman web dengan didasarkan design template tertentu maka web anda akan memiliki gaya halaman web yang sama dan terlihat konsisten dari halaman perhalaman.
3. Kemudahan dan efisiensi dalam penggunaan.

   Program ini tidak hanya dirancang untuk anda yang sudah mahir dan mengerti bahasa pemrograma. Bagi anda yang belum mengerti bahasa pemrograman, anda bisa membuat halaman web dengan hanya cara mengklik atau drag and drop menggunakan mouse serta anda juga bisa melihat halaman html-nya selama proses desain berlangsung.
4. Mudah untuk mengupload melalui FTP.

   Dreamweaver sudah dilengkapi dengan fitur FTP jadi setelah anda selesai membangun sebuah web, anda bisa langsung menguploadnya melalui FTP . FTP(singkatan dari *File Transfer Protocol*) adalah sebuah protokol Internet yang berjalan di dalam lapisan aplikasi yang merupakan standar untuk pentransferan berkas (*file*) komputer antar mesin-mesin dalam sebuah *internetwork*.
5. Dapat dikustom

   Dreamweaver dapat disesuaikan dengan kebutuhan yang anda perlukan. Menu, tab, perintah, font dan warna semua kode dapat disesuaikan

dengan preferensi pribadi. Hal ini dapat secara efektif memudahkan proses desain web. Selain itu dreamweaver didukung banyak plug-in yang membantu anda dalam proses desain.