BAB 2
LANDASAN TEORI

2.1 Pengertian Novel

Novel berasal dari bahasa Itali, juga dari bahasa Latin yakni novellus yang diturunkan pula dari kata novies yang berarti baru. Dikatakan baru karena kalau dibandingkan dengan jenis-jenis sastra lainnya seperti puisi, drama, dan lain-lain, maka jenis novel ini muncul kemudian.

Novel adalah karangan yang panjang dan berbentuk prosa dan mengandung rangkaian cerita kehidupan seseorang dengan orang lain di sekelilingnya dengan menonjolkan watak dan sifat setiap pelaku.

Berdasarkan beberapa pendapat di atas, penulis menyimpulkan bahwa novel adalah suatu cerita fiksi yang menggambarkan kisah hidup tokoh melalui rangkaian peristiwa yang kompleks dan mengubah nasib tokoh tersebut [5].

2.2 Ekstraksi Informasi

Ekstraksi Informasi adalah pengambilan fakta dan informasi terstruktur dari isi koleksi teks yang besar. Pengertian fakta disini adalah beragam entitas yang diperhitungkan. Secara singkat ekstraksi informasi adalah sebagai sebuah proses untuk menemukan informasi terstruktur dari dokumen tidak terstruktur atau semi terstruktur [6].

Ekstraksi informasi dapat berguna dalam beberapa bidang seperti pencarian lowongan pekerjaan, pengumuman seminar, dan iklan penjualan. Ekstraksi informasi dari data teks dengan metode machine learning juga dapat digunakan untuk membangun sebuah database yang berisi informasi sehingga data teks yang tersedia online dapat dicari dengan mudah. Contohnya adalah iklan-iklan lowongan pekerjaan yang dapat diekstraksi menjadi database lowongan pekerjaan yang lebih terstruktur.

Dalam makalah ilmiah, ekstraksi informasi dapat berguna untuk membangun database makalah yang dapat dicari berdasarkan data penulis atau

afiliasinya. Hal tersebut memudahkan orang untuk mencari kontak dari penulis makalah dan melihat perkembangan makalah yang dibuat oleh penulis tertentu.

Berdasarkan penjelasan dari kutipan di atas, dapat disimpulkan bahwa ekstraksi informasi adalah proses penyaringan parameter-parameter informasi fakta dari data yang telah tersedia.

2.3 Preprocessing

Preprocessing adalah tahapan untuk mempersiapkan teks menjadi data yang akan diolah di tahapan berikutnya. Input-an awal pada proses ini adalah berupa dokumen asli dan dokumen yang akan di uji. Preprocessing pada penelitian ini cleansing untuk membersih data yaitu simbol dan kata yang yang tidak diperlukan pada hasil ekstraksi informasi

2.4 Ekstraksi Fitur

Fitur digunakan sebagai masukan dalam metode rule-based sistem. Ekstraksi fitur perlu dilakukan dalam ekstraksi informasi, karena data teks yang bentuknya tidak terstruktur tidak memiliki fitur, sehingga tidak bisa langsung dimasukkan kedalam metode *rule-based* sistem. Menurut penelitian pada [7], ada tiga jenis fitur yang dapat digunakan untuk ekstraksi informasi, yaitu:

1. Fitur lokal, yaitu ciri-ciri khusus yang ada di dalam masing-masing kalimat.
2. Fitur tata letak, yaitu posisi kalimat dalam barisnya masing-masing.
3. Fitur lexicon, yaitu jenis kata kunci yang terkandung dalam kalimat tersebut.

Menurut penelitian dari Peng, fitur tata letak dapat secara drastis meningkatkan kinerja dari fitur lokal [6]. Sedangkan fitur lexicon peningkatan kinerjanya tidak sebesar fitur tata letak. Bahkan penambahan kombinasi fitur tata letak dan lexicon dapat menurunkan kinerjanya.

2.5 Pendekatan Sistem Ektraksi Informasi

Dalam pembangunan SEI terdapat dua pendekatan dasar, yaltu pendekatan hand-crafing atau disebut juga knowledge engineering dan pendekatan machine learning (Bosch & Daelemans, 2005). Kedua pendekatan tersebut memiliki kelebihan dan kekurangannya masing-masing.

2.6 Rule-based Sistem

rule base dikenal juga dengan sebutan pendekatan hand- crafting atau Knowledge engineering. Disebut pendekatan rule base karena pendekatan ini memanfaatkan keahlian manusia (human expert) untuk membuat aturan-aturan (rules) secara manual melalui proses pemahaman domain. Human expert atau pakar dituntut untuk bisa memahami SEI dan domain sistem. Jika domain sistem terlalu besar maka dibutuhkan sumber daya manusia yang cukup banyak.

Setelah pakar membuat aturan, sistem diuji coba pada sekumpulan dokumen. Keluaran dari proses uji coba tersebut diperiksa untuk melihat apakah aturan-aturan yang dlbuat sudah mencukupi atau belum. Jika aturan yang dlbuat belum memberikan kinerja sistem yang maksimal maka aturan-aturan tersebut dimodifikasi dan proses pengujian kembali dilakukan.

Sistem berbasis aturan (*rule-based sistem*) adalah jalan untuk menyimpan atau memanipulasi pengetahuan untuk menintrepentasikan informasi yang berguna [3]. Biasanya sistem ini diimplementasikan dengan aplikasi Artificial Intelligence. Suatu aturan terdiri dari 2 bagian, yaitu:

1. Antacedent, yaitu bagian yang mengekspresikan situasi atau premis (Pernyataan berawalan IF).
2. Consequent, yaitu bagian yang menyatakan suatu tindakan tertentu atau konklusi yang diterapkan jika situasi atau premis bernilai benar (Pernyataan berawalan THEN).

Umumnya, sebuah aturan dapat mempunyai gabungan beberapa antecedent dengan kata kunci AND (konjungsi), OR (disjungsi), atau kombinasi keduanya. Metode Aturan dapat dilihat pada ilustrasi berikut.

```
 IF <antecedent 1>
AND <antecedent 2>
        .
        .
        .
AND <antecedent n>
 THEN <consequent>
 IF <antecedent 1>
 OR <antecedent 2>
        .
        .
```

.

```
OR <antecedent n>
THEN <consequent>
```

Gambar 2. 1 Contoh Algoritma Rule-Based Antacedent dan Consequent

Kelebihan pendekatan rule base adalah pencapaian kinerja sistem yang tinggi lebih mudah untuk dilakukan karena penggunaan keahlian manusia. Pendekatan ini juga tidak terlalu sulit selama terdapat pakar yang memahami domain sistem dengan baik. Di sisi lain, hal ini juga sekaligus menjadi kelemahannya karena sistem menjadi sangat tergantung pada keberadaan pakar. Padahal untuk mendapatkan satu orang pakar saja apalagi beberapa pakar adalah hal yang tidak mudah.

Kekurangan lain dari pendekatan ini adalah diperlukan waktu yang lama dan biaya yang tinggi. Biaya yang tinggi ini disebabkan kebutuhan terhadap sumber daya manusia yang banyak terlebih jika sistem memillki ruang Imgkup yang sangat besar. Sistem yang dikembangkan dengan dengan pendekatan rule base juga akan mengalami masalah adaptability, yaitu ketika pakar yang membuat aturan-aturan dalam sistem sudah tidak ada sehingga pakar yang baru sulit untuk melakukan penyesuaian jika ingin melakukan perubahan pada

2.7 Informasi tersturktur

Terstruktur pada kamus bahasa indonesia adalah sudah dalam keadaan disusun dan diatur rapi. Sedangkan informasi pada kamus bahasa indonesia adalah pengumpulan atau data yang diolah sehingga dapat dijadikan dasar untuk mengambil keputusan yang tepat. Pengertian informasi tersturktur ialah suatu kalimat atau teks yang telah mempunyai aturan dan pola yang tersusun rapih

2.8 Converter

Convert adalah perintah untuk mengubah suatu file menjadi format yang berbeda agar dapat dibaca oleh aplikasi yang bersangkutan.

2.8.1 OCR (Optical Character Recognition)

OCR (Optical Character Recognition) adalah aplikasi yang berfungsi untuk men scan gambar pada image dan dijadikan text, dan aplikasi ini juga bisa

menjadi support /aplikasi tambahan untuk scanner. Dengan adanya **OCR**, Image yang bertulisan tangan, tulisan mesin ketik atau computer text, dapat dimanipulasi[13].

2.9 Data Flow Diagram (DFD

Data Flow Diagram (DFD) merupakan diagram yang menggunakan notasi notasi atau symbol simbol untuk mengambarkan sistem jaringan kerja antar fungsi-fungsi yang berhubungan satu sama lain dengan aliran dan penyimpanan data.

DFD sering digunakan untuk menggambarkan suatu sistem yang telah ada atau sistem baru yang akan dikembangkan secara logika tanpa mempertimbangkan lingkungan fisik dimana data tersebut mengalir atau dimana data tersebut akan disimpan. Salah satu keuntungan menggunakan diagram aliran data adalah memudahkan pemakai (user) yang kurang menguasai bidang computer untuk mengerti sistem yang akan dikerjakan.

DFD terdiri dari diagram konteks (context diagram) dan diagram rinci (level diagram). Diagram konteks adalah diagram yang terdiri dari suatu proses dan menggambarkan ruang lingkup suatu sistem. Diagram konteks merupakan level tertinggi dari DFD yang menggambarkan seluruh input ke sistem atau output dari sistem. Dalam diagram konteks biasanya hanya ada satu proses. Tidak boleh ada store dalam diagram konteks. Diagram rinci adalah diagram yang menguraikan proses apa yang ada dalam diagram level di atasnya.

2.10 Kamus Data

Kamus data atau data dictionary adalah katalog fakta tentang data dan kebutuhan-kebutuhan informasi dari suatu sistem informasi. Dengan kamus data analis sistem dapat mendefinisikan data yang mengalir di sistem dengan lengkap. Pada tahap analisis sistem, kamus data digunakan sebagai alat komunikasi antara analis sitem dengan pemakai sistem tentang data yang mengalir ke sistem, yaitu tentang data yang masuk ke sistem dan tentang informasi yang dibutuhkan oleh pemakai sistem. Pada tahap perancangan sistem, kamus data digunakan untuk merancang input, merancang laporan-laporan dan database.

Kamus data dibuat berdasarkan arus data yang ada di Data Flow Diagram (DFD). Arus data di DFD sifatnya adalah global, hanya di tunjukkan nama arus datanya saja. Keterangan lebih lanjut tentang struktur dari suatu arus data di DFD secara lebih terinci dapat di lihat di kamus data. DD tidak menggunakan notasi grafik sebagaimana halnya DFD. DD berfungsi membantu pelaku sistem untuk mengerti aplikasi secara detil, dan mereorganisasi semua elemen data yang digunakan dalam sistem secara presisi sehingga pemakai dan penganalisa sistem punya dasar pengertian yang sama tentang masukan, keluaran, penyimpanan dan proses. DD mendefinisikan elemen data dengan fungsi sebagai berikut:

1. Menjelaskan arti aliran data dan penyimpanan dalam DFD
2. Mendeskripsikan komposisi paket data yang bergerak melalui aliran, misalnya alamat di uraikan menjadi kota, kodepos, propinsi, dan negara.
3. Mendeskripsikan komposisi penyimpanan data.
4. Menspesifikasikan nilai dan satuan yang relevan bagi penyimpanan dan aliran.
5. Mendeskripsikan hubungan detil antara penyimpanan yang akan menjadi titik perhatian dalam entity relationship diagram.

2.11 ERD (Entity Relationship Diagram)

ERD merupakan suatu model untuk menjelaskan hubungan antar data dalam basis data berdasarkan objek-objek dasar data yang mempunyai hubungan antar relasi. ERD untuk memodelkan struktur data dan hubungan antar data, untuk menggambarkannya digunakan beberapa notasi dan simbol. Pada dasarnya ada tiga simbol yang digunakan, yaitu :

a. Entitas

Entiti merupakan objek yang mewakili sesuatu yang nyata dan dapat dibedakan dari sesuatu yang lain. Simbol dari entiti ini biasanya digambarkan dengan persegi panjang.

b. Atribut

Setiap entitas pasti mempunyai elemen yang disebut atribut yang berfungsi untuk mendeskripsikan karakteristik dari entitas tersebut. Isi dari atribut

mempunyai sesuatu yang dapat mengidentifikasikan isi elemen satu dengan yang lain. Gambar atribut diwakili oleh simbol elips.

c. Hubungan / Relasi

Hubungan antara sejumlah entitas yang berasal dari himpunan entitas yang berbeda. Relasi dapat digambarkan sebagai berikut :

Relasi yang terjadi diantara dua himpunan entitas (misalnya A dan B) dalam satu basis data yaitu:

1. Satu ke satu (*One to one*)

Hubungan relasi satu ke satu yaitu setiap entitas pada himpunan entitas A berhubungan paling banyak dengan satu entitas pada himpunan entitas B.

2. Satu ke banyak (*One to many*)

Setiap entitas pada himpunan entitas A dapat berhubungan dengan banyak entitas pada himpunan entitas B, tetapi setiap entitas pada entitas B dapat berhubungan dengan satu entitas pada himpunan entitas A.

3. Banyak ke banyak (*Many to many*)

Setiap entitas pada himpunan entitas A dapat berhubungan dengan banyak entitas pada himpunan entitas B

2.12 Bahasa Pemograman

Bahasa pemrograman merupakan kumpulan aturan yang disusun sedemikian rupa sehingga memungkinkan pengguna komputer membuat program yang dapat dijalankan dengan aturan tersebut [8]. Beberapa contoh bahasa pemrograman yang banyak dipakai, yaitu C, C++, C#, Pascal, Java, JavaScript, PHP, SQL, Python dan . Bahasa Pemrograman yang digunakan pada pembangunan sumber daya jenis kata ini adalah menggunakan PHP, JavaScript dan SQL.

2.13 PHP

Hypertext Prepocessor (PHP) adalah skrip yang menyatu dengan kode HTML dan berjalan pada sisi server (*server-side*). Artinya semua sintaks yang kita berikan akan sepenuhnya dijalankan pada server secangkan yang dikirimkan kebrowser hanya hasilnya [9]. Keuntungan menggunakan PHP adalah tidak diedarkannya kode sumber yang membangun PHP ke sisi klien sehingga

kerahasiaan kode dapat dilindungi. PHP berjalan pada sisi server, sehingga untuk dapat menggunakannya kita harus mengaktifkan web server terlebih dahulu baik offline maupun online.

Deklarasi variabel dengan tipe data tidak diperlukan dalam PHP seperti bahasa pemrograman pascal. Setiap variabel yang terbentuk dalam program dianggap bertipe *variant*, dengan kata lain dapat menampung tipe data dengan jenis apapun (*dynamic variable*). Sistem yang dibangun menggunakan PHP versi 5.6.19

PHP memiliki keunggulan seperti berikut:

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, Oracle, Microsoft SQL Server, Interbase, dan banyak lagi.
4. PHP bersifat open source dan gratis.

2.13.1 HTML (Hyper Text Markup Language)

HyperText Markup Language, sering disebut sebagai HTML, merupakan standar bahasa markup yang digunakan untuk membuat halaman web. Web browser dapat membaca file HTML dan membuat mereka ke dalam halaman web terlihat atau terdengar. HTML menggambarkan struktur dari situs semantik bersama dengan isyarat untuk presentasi, membuatnya menjadi bahasa markup, daripada bahasa pemrograman.

Elemen HTML membentuk blok bangunan dari semua website. HTML memungkinkan gambar dan objek yang akan tertanam dan dapat digunakan untuk membuat bentuk-bentuk interaktif. Menyediakan sarana untuk membuat dokumen terstruktur oleh struktural menunjukkan semantik untuk teks seperti judul, paragraf, daftar, link, kutipan dan item lainnya.

Bahasa ini ditulis dalam bentuk elemen HTML terdiri dari tag diapit kurung sudut (seperti <html>). Browser tidak menampilkan tag HTML dan script, tetapi menggunakannya untuk menafsirkan isi halaman.

HTML dapat menanamkan script yang ditulis dalam bahasa seperti JavaScript yang mempengaruhi perilaku halaman web HTML. Web browser juga dapat merujuk ke Cascading Style Sheets (CSS) untuk menentukan tampilan dan tata letak teks dan bahan lainnya. The World Wide Web Consortium (W3C), pengelola dari kedua HTML dan CSS standar, mendorong penggunaan CSS lebih eksplisit HTML presentasi sejak tahun 1997.

2.13.2 JavaScript (AJAX & jQuery)

HTTP (*Hypertext Transfer protocol*) merupakan protokol utama dibalik WWW (*World Wide Web*). Protokol ini yang menangani semua permintaan halaman website, gambar, link, dan form. Dengan adanya HTTP memungkinkan sebuah cara yang terstandarisasi untuk komputer berkomunikasi dengan computer lainnya. HTTP akan menspesifikasikan bagaimana klien meminta data, dan bagaimana server memproses permintaan tersebut. HTTP memiliki tiga tipe permintaan dasar : POST, GET, dan HEAD [12]. Permintaan HEAD digunakan untuk memverifikasi bahwa sebuah halaman web yang diminta masih tersedia. Sebuah mesin pencari biasanya menggunakan permintaan HEAD untuk memastikan bahwa hasil pencariannya masih aktif. Jika permintaan HEAD gagal, maka mesin pencari tersebut bisa langsung menghapus hasil pencarian dari database yang dimiliki oleh mesin pencari tersebut.

Javascript merupakan salah satu bahasa pemrograman yang pertama kali dikembangkan oleh Netscape yang digunakan dalam sebuah browser atau perambah web [10]. Javascript dibuat agar mudah diintegrasikan ke dalam program dan aplikasi berbasiskan web. AJAX merupakan singkatan dari Asynchronous Javascript and XML merupakan bagian dari javascript yang bertugas atau berfungsi untuk memodifikasi sebagian atau semua dari sebuah halaman web secara dinamis pada saat dibutuhkan saja [10].

2.14 Pengelola Database

Database merupakan kumpulan file-file yang berhubungan secara logis dan digunakan secara rutin pada operasi-operasi sistem informasi manajemen. Semua database umumnya berisi elemen-elemen data yang disusun ke dalam file-file yang diorganisasikan berdasarkan sebuah skema atau struktur tertentu, tersimpan di hardware komputer dan dengan software untuk melakukan manipulasi data untuk kegunaan tertentu. Jadi, suatu database adalah menunjukkan suatu kumpulan tabel yang dipakai dalam suatu lingkup perusahaan atau instansi untuk tujuan tertentu [12].

Software yang digunakan untuk mengelola dan permintaan panggilan (*query*) basis data yang disebut sistem manajemen database (*database management syste*m, DBMS).

Fungsi dasar database:

1. Pengelompokan data, database bertujuan untuk mengklasifikasikan data agar mudah dipahami. Misalnya dalam sistem perpustakaan, ada kelompok data buku, penerbit, transaksi peminjaman, dan mahasiswa.
2. Menghindari duplikasi dan inkonsistensi dalam data.
3. Membuatnya mudah untuk menyimpan, mengakses, dan update, dan menghapus data.
4. Menjamin kualitas data dan informasi yang dapat diakses sesuai dengan dimasukkan (integritas data)
5. Sebuah solusi dalam proses penyimpanan data, terutama data.
6. Mendukung kinerja aplikasi yang memerlukan penyimpanan data.

2.14.1 MySQL

MySQL merupakan database server yang bersifat multiuser dan multi threaded. SQL adalah bahasa database standar yang memudahkan penyimpanan, pengubahan dan akses informasi. Pada MySQL dikenal dengan istilah database dan tabel. Tabel sendiri merupakan sebuah struktur data dua dimensi yang terdiri dari baris-baris record dan kolom [11]. MySQL termasuk salah satu database open source yang paling banyak digunakan karena selain gratis MySQL pun

mempunyai banyak dukungan bahasa pemrograman dan aplikasi sebagai solusi database.

MySQL memiliki keunggulan sebagai berikut:

1. Portabilitas. MySQL dapat berjalan stabil pada berbagai sistem operasi, seperti Windows, Linux, FreeBSD, Mac Os Server, dan masih banyak lagi.
2. *Open Source*. MySQL didistribusika GPL sehingga dapat digunakan secara cuma-cuma.
3. *Multiuser*. MySQL dapat digunakan oleh beberapa pengguna dalam waktu bersamaan tanpa mengalami masalah konflik
4. *Performance Tuning*. MySQL mampu memproses lebih banyak query SQL per satuan waktu.
5. Jenis Kolom. MySQL memiliki tipe kolom yang sangat kompleks, seperti integer, float, double,char, text, date, datetime, dan lain lain.

MySQL dapat digunakan untuk membuat dan mengola database beserta isinya. Kita dapat memanfaatkan MySQL untuk menambahkan, mengubah dan menghapus data yang berada dalam database. MySQL merupakan sisitem manajemen database yang bersifat at relational. Artinya data-data yang dikelola dalam database akan diletakkan pada beberapa tabel yang terpisah sehingga manipulasi data akan menjadi jauh lebih cepat.

2.14.2 SQL (Structured Query Language)

Structured Query Language adalah bahasa yang digunakan untuk mengelola data pada RDBMS. SQL digunakan untuk mengambil, menambahkan, dan memanipulasi data dari sebuah basis data relasional [11].

Bahasa SQL terbagi menjadi dua kategori yaitu kategori pertama schema statements kategori pernyataan (statements) untuk membuat objek basis data berupa tabel, index, constraints, dan sebagainya. Kategori kedua (SQL data statements) yaitu pernyataan yang digunakan untuk memanipulasi dan mengambil data yang tersimpan dalam basis data.

2.15 Perangkat Lunak Pendukung

Perangkat lunak adalah istilah khusus untuk data yang diformat, dan disimpan secara digital, termasuk program komputer, dokumentasinya, dan berbagai informasi yang bisa dibaca, dan ditulis oleh komputer. Dengan kata lain, bagian sistem komputer yang tidak berwujud.

Dalam peneliti ini menggunakan WAMP sebagai web server untuk pembangunan dan pengujian kamus jenis kata yang dibangun, WebBrowser, dengan notepad++ sebagai penyunting kode bahasa pemrograman.

2.16 Pengujian

Pengujian adalah suatu proses mengevaluasi suatu sistem atau komponen dengan maksud untuk menemukan apakah itu memenuhi persyaratan yang ditentukan atau tidak. Dengan kata lain pengujian ialah mengeksekusi sistem untuk mengidentifikasi kesenjangan, atau kesalahan.

2.16.1 Web Browser

Web Browser adalah layanan internet untuk menjelajahi atau berselancar di dunia maya dengan menggunakan jaringan internet [12]. Fungsi dari Web Browser sendiri adalah untuk menampilkan halaman web atau melakukan interaksi dengan dokumen yang disediakan server. Dalam penelitian ini penulis menggunakan mozilla firefox versi 46.0.1

2.16.2 Notepad++

Notepad++ adalah program aplikasi yang berguna untuk mengedit tes dan skrip kode pemrograman seperti HTML, CSS, PHP, XML, Java, dan lain-lain yang bekerja pada sistem operasi windows. Kelebihan Notepad++ jika dibandingkan dengan notepad bawaan windows adalah memiliki kelengkapan fitur untuk mempermudah pengguna saat mengedit kode termasuk saat mengedit kode HTML dan kode CSS.