

## **BAB 2**

### **LANDASAN TEORI**

#### **2.1 *Natural Language Processing***

*Natural Language Processing* (NLP) adalah bidang ilmu dan teknik computer yang telah berkembang dari studi bahasa dan komputasi linguistic dalam bidang kecerdasan buatan. Tujuan dari NLP adalah untuk mendesain dan membangun aplikasi yang dapat memfasilitasi interaksi antar manusia dengan mesin atau perangkat lain dengan menggunakan bahasa alami manusia [6]. Salah satu bidang dalam NLP adalah *part-of-speech tagging* yang merupakan salah satu pembahasan pada penelitian ini.

#### **2.2 Visualisasi**

Menurut Card dan Mackinlay, definisi visualisasi adalah menggunakan teknologi komputer sebagai pendukung untuk melakukan penggambaran data visual yang interaktif untuk memperkuat pengamatan. Sedangkan menurut McCormick, visualisasi adalah metode penggunaan komputer untuk mentransformasikan simbol menjadi geometrik dan memungkinkan peneliti dalam hal mengamati simulasi komputasi yang dapat memperkaya proses penemuan ilmiah sehingga dapat mengembangkan pemahaman yang lebih dalam dan tak terduga.

Berdasarkan kedua pernyataan diatas dapat disimpulkan bahwa visualisasi adalah suatu teknik penggunaan komputer untuk menentukan metode terbaik dalam menampilkan data. Dengan menggunakan visualisasi, data yang ditampilkan dapat mempermudah pengguna untuk melihat data yang sulit dilihat dan agar dapat mengembangkan pemahaman yang lebih dalam. Beberapa contoh visualisasi adalah dengan menampilkan data atau informasi dalam bentuk gambar, grafik, struktur hirarki, pola, dan warna. Pada penelitian ini, visualisasi digunakan untuk menggambarkan data berupa teks yang telah dianotasi atau diberi label.

### 2.3 User Experience

Menurut definisi dari ISO 9241-210, *user experience* adalah persepsi atau pengalaman seseorang dan responnya dari penggunaan sebuah produk, sistem, atau jasa. *User experience* menilai seberapa kepuasan dan kenyamanan seseorang terhadap sebuah produk, sistem, dan jasa. Sedangkan menurut Jerry Cao, dkk., *user experience* adalah istilah umum untuk keseluruhan pengalaman dari pengguna terhadap sebuah produk. Seperti misalnya apa yang mereka sukai, betapa mudahnya mereka mencapai tujuan mereka, apa saat-saat menyenangkan dan membuat frustrasi ketika menggunakan produk [7].

Berdasarkan kedua pernyataan di atas dapat disimpulkan bahwa *user experience* adalah pengalaman yang dirasakan seseorang dalam menggunakan sebuah produk yang dapat menggambarkan penilaian tentang tingkat kenyamanan dan kemudahan yang diberikan produk tersebut.

### 2.4 Tag Kelas Kata atau (*Part-of-Speech Tag*)

*Part-of-Speech (POS) tag* adalah suatu proses pemberian label setiap kata atau token berdasarkan kelas kata (*tagset*) yang sesuai. *POS tag* adalah tahapan yang sangat penting dan sering digunakan dalam pemrosesan NLP. Informasi nilai *POS tag* ini merupakan hal yang mendasar bagi keperluan pengolahan bahasa manusia. Beberapa metode pendekatan yang sering digunakan dalam melakukan *POS tag* adalah berdasar aturan (*rule based*), pendekatan probabilistik, dan pendekatan berbasis transformasi (*transformational based*). *Tagset* dalam Bahasa Indonesia memiliki beberapa variasi [8] [9] [10]. Dalam penelitian ini, *tagset* yang digunakan dapat dilihat di Tabel 2.1.

**Tabel 2.1 Tagset Bahasa Indonesia [10].**

Tag	Deskripsi	Contoh
CC	Konjungtor koordinatif	dan, tetapi, atau
CD	Numeralia kardinal	dua, juta, enam, 7916, sepertiga, 0,025, 0,525, banyak, kedua, ribuan, 2007, 25
OD	Numeralia ordinal	ketiga, ke-4, pertama
DT	Artikel	Para, Sang, Si

Tag	Deskripsi	Contoh
FW	Kata bahasa asing	<i>climate change, terms and conditions</i>
IN	Preposisi	dalam, dengan, di, ke, oleh, pada, untuk
JJ	Adjektiva	bersih, panjang, hitam, lama, jauh, marah, suram, nasional, bulat
MD	Verba modal dan verba bantu	boleh, harus, sudah, mesti, perlu
NEG	Kata ingkar	tidak, belum, jangan
NN	Nomina	monyet, bawah, sekarang, rupiah
NNP	<i>Proper noun</i>	Boediono, Laut Jawa, Indonesia, India, Malaysia, Bank Mandiri, BBKP, Januari, Senin, Idul Fitri, Piala Dunia, Liga Primer, <i>Lord of the Rings: The Return of the King</i>
NND	Penggolong atau nomina ukuran	orang, ton, helai, lembar
PR	Pronomina penunjuk	ini, itu, sini, situ
PRP	Pronomina persona	saya, kami, kita, kamu, kalian, dia, mereka
RB	Adverbia	sangat, hanya, justru, niscaya, segera
RP	Partikel	pun, -lah, -kah
SC	Konjungtor subordinatif	sejak, jika, seandainya, supaya, meski, seolah-olah, sebab, maka, tanpa, dengan, bahwa, yang, lebih ... daripada ..., semoga
SYM	Simbol	IDR, +, %, @
UH	Interjeksi	brengsek, oh, ooh, aduh, ayo, mari, hai
VB	Verba	merancang, mengatur, pergi, bekerja, tertidur
WH	Pronomina penanya	siapa, apa, mana, kenapa, kapan, di mana, bagaimana, berapa
X	Tidak / belum diketahui	statemen
Z	Tanda baca	"...", ?, .

## 2.5 Basis Data

Basis data adalah kumpulan data yang saling berhubungan secara logikal serta deskripsi dari data tersebut, yang dirancang untuk memenuhi kebutuhan

informasi suatu organisasi. Basis Data adalah sebuah penyimpanan data yang besar yang bisa digunakan oleh banyak pengguna. Semua data terintegrasi dengan jumlah duplikasi yang minimum. Basis Data tidak lagi dipegang oleh satu departemen, tetapi dibagikan ke seluruh departemen pada perusahaan. Basis Data itu sendiri tidak hanya memegang data operasional organisasi tetapi juga penggambaran dari data tersebut [11].

## 2.6 *Annotated Corpus*

*Corpus* merupakan sekumpulan teks yang dapat dibaca oleh mesin dan dibuat dengan pengaturan komunikatif alami. Biasanya *corpus* dijadikan sampel representatif dan dikelompokkan berdasarkan faktor tertentu, misalnya artikel berita, blog, fiksi literatur, catatan, dan lain sebagainya. Sekumpulan *corpus* yang sudah dianotasi disebut *annotated corpus*. *Annotated corpus* biasanya digunakan sebagai *data training* untuk algoritma *machine learning* [6]. *Annotated corpus* merupakan data keluaran yang dihasilkan dalam penelitian ini.

## 2.7 MySQL

Menurut Bimo Sunarfrihantono, MySQL merupakan *multi-user database* yang menggunakan bahasa *Structured Query Language (SQL)* [12]. Sedangkan menurut Lukmanul Hakim dan Uus Musalini, MySQL merupakan salah satu jenis program *database server*, yaitu sebuah program yang berfungsi untuk mengolah, menyimpan dan memanipulasi data dari server [13].

Berdasarkan definisi MySQL dari pernyataan diatas dapat disimpulkan bahwa MySQL adalah sebuah *software database* yang dijalankan menggunakan perintah SQL untuk mengolah, menyimpan, dan memanipulasi data sesuai dengan keinginan pengguna. Pada penelitian ini, MySQL digunakan untuk menyimpan data teks, token yang telah dianotasi, *tag* kelas kata, dan pengguna.

## 2.8 PHP

Menurut Didik Dwi Prasetyo, PHP merupakan bahasa *scripting server-side*, di mana pemrosesan datanya dilakukan pada sisi server. Sederhananya, server lah yang akan menerjemahkan skrip program, baru kemudian hasilnya akan dikirim

kepada *client* yang melakukan permintaan [13]. Sedangkan menurut Badiyanto, PHP: *Hypertext Preprocessor* adalah bahasa pemrograman yang dapat ditanamkan atau disisipkan ke dalam HTML atau PHP banyak dipakai untuk membuat situs web dinamis. Sebagian besar sintaksis mirip dengan bahasa C, Java, dan Perl, ditambah beberapa fungsi PHP yang spesifik. Tujuan utama penggunaan bahasa ini adalah untuk memungkinkan perancangan web dinamis [14].

Berdasarkan definisi PHP dari kedua pernyataan di atas, dapat disimpulkan bahwa PHP adalah bahasa pemrograman yang dapat digunakan untuk membuat sebuah web dinamis yang berjalan pada sisi server (*server-side*), sehingga semua kode program dijalankan pada sisi server dan hasilnya akan dikirim kepada klien melalui *browser*. Pada penelitian ini, PHP digunakan sebagai bahasa pemrograman sisi server.

## **2.9 Javascript**

Menurut Edy Winarno, dkk javascript adalah bahasa *scripting* seperti HTML yang artinya kode langsung dieksekusi dan tidak dikompilasi terlebih dahulu untuk dijadikan *file executable* [15]. Javascript disisipkan ke halaman HTML dan pengeksesusiannya berdasarkan *client side* yang artinya kode dieksekusi di sisi klien (sisi browser) bukan di sisi server. Pada penelitian ini, javascript digunakan membuat halaman web lebih interaktif dengan memberikan efek (animasi) atau memanipulasi DOM (*Document Object Model*) HTML.

