

BAB 2

LANDASAN TEORI

2.1 Database

Database adalah sekumpulan data yang saling berhubungan secara logikal, serta deskripsi dari data yang akan dirancang untuk menghasilkan informasi yang dibutuhkan dalam sebuah organisasi[6]. Sedangkan sistem database merupakan kumpulan program aplikasi yang berinteraksi dengan database bersama dengan DBMS dan database itu sendiri.

Maka lingkup terbesar dalam suatu organisasi data adalah sistem database. Sistem database ini mencakup semua bentuk komponen data yang ada dalam sistem dimana database merupakan komponen penyusun utamanya.

2.2 Structured Query Language (SQL)

Structured Query Language atau dikenal dengan SQL merupakan suatu bahasa (language) yang digunakan untuk mengakses database. SQL adalah suatu bahasa yang terstruktur[7]. ISO dan ANSI telah ditetapkan sebagai standarisasi bahasa SQL. Standarisasi ini tidak bergantung pada mesin yang di gunakan dan hampir semua softwarenya dapat mengerti dan mengenal bahasa SQL ini[8]. Perintah SQL disebut juga query. Ada tiga sub perintah padaperintah SQL, yaitu *Data Definition Language*, *Data Manipulation Language*, dan *Data Control Language*.

2.3 Data Definition Language (DDL)

Data Definition Language atau DDL merupan sub bahasa SQL yang berfungsi untuk membangun kerangka database. Terdapat tiga perintah yang termasuk kedalam perintah DDL, yaitu :

a. CREATE

Perintah CREATE ini digunakan untuk membuat database baru, tabel baru, view baru, dan membuat kolom.

```
CREATE DATABASE [IF NOT EXISTS] nama_database;
```

b. RENAME

Perintah RENAME ini digunakan untuk mengubah nama suatu tabel.

```
RENAME TO nama_tabel_baru
```

c. ALTER

Perintah ALTER ini digunakan untuk membuat struktur dari tabel yang sudah dibuat. Hal yang dapat dilakukan yaitu mengganti nama tabel, menambah kolom, mengubah kolom, menghapus kolom, serta memberi atribut pada kolom.

```
ALTER TABLE nama_tabel alter_options;
```

d. DROP

Perintah DROP ini digunakan untuk menghapus database ataupun menghapus tabel.

```
DROP DATABASE [IF EXISTS] nama_database;
```

2.4 *Data Manipulation Language (DML)*

Data Manipulation Language atau DML merupakan sub bahasa SQL yang berfungsi untuk memanipulasi data yang terdapat pada database yang telah dibuat. Terdapat empat perintah yang termasuk kedalam perintah DML, yaitu :

a. INSERT

Perintah INSERT dapat digunakan setelah pembuatan database dan tabel. Perintah ini berfungsi untuk memanipulasi data dengan memasukkan ataupun menyisipkan data baru kedalam tabel.

```
INSERT INTO nama_tabel VALUES ('nilai1','nilai2',...);
```

Atau

```
INSERT INTO nama_tabel(field1,field2,...)  
VALUES ('nilai1','nilai2',...);
```

Atau

```
INSERT INTO nama_tabel  
SET field1='nilai1', field2='nilai2',...;
```

b. SELECT

Perintah SELECT ini berfungsi untuk mengambil atau menampilkan data dari suatu tabel ataupun beberapa tabel yang saling berelasi.

```
SELECT [field | *] FROM nama_tabel [WHERE kondisi];
```

c. UPDATE

Perintah UPDATE dapat digunakan jika terdapat data yang salah atau perlu diperbaiki. Perintah ini berfungsi untuk memperbaiki data yang sudah ada.

```
UPDATE nama_tabel SET field1='nilaibaru'  
[WHERE kondisi];
```

d. DELETE

Perintah DELETE dapat digunakan jika ingin menghapus data yang ada pada tabel. Perintah ini tidak dapat digagalkan apabila perintah telah di jalankan sehingga data tidak dapat dikembalikan.

```
DELETE FROM nama_tabel [WHERE kondisi];
```

2.5 Data Control Language (DCL)

Data Control Language atau DCL merupakan sub bahasa SQL yang berfungsi untuk mengontrol data dan server databasenya. Terdapat dua perintah yang termasuk kedalam perintah DCL, yaitu :

a. GRANT

Perintah GRANT dapat digunakan oleh administrator server untuk memberikan hak akses kepada user. Hak akses yang dapat diberikan berupa CREATE, SELECT, DELETE, UPDATE, dan hak-hak lainnya yang berkaitan dengan database.

```
GRANT priv_type  
ON {tbl_name | * | *.* | db_name.*}  
TO user_name [IDENTIFIED BY 'password']  
[WITH GRANT OPTION]
```

b. REVOKE

Perintah REVOKE dapat digunakan oleh administrator server untuk mencabut hak akses user. Perintah ini merupakan kebalikan dari perintah GRAND. Hak akses yang dapat dicabutpun sama yaitu CREATE, SELECT, DELETE, UPDATE, dan hak-hak lainnya yang berkaitan dengan database.

```
REVOKE priv_type
ON {tbl_name | * | *.* | db_name.*}
FROM user_name
```

2.6 Subquery

Subquery adalah dimana terdapat suatu query yang berada dalam query lainnya. Subquery ini biasa dipakai untuk menangani masalah yang kompleks dan akan sulit jika hanya menggunakan sebuah query saja. Subquery juga dapat digunakan sebagai alternatif pengganti dari penggunaan join dan union. Terdapat tiga jenis *subquery*, yaitu :

a. *Scalar subquery*

Jenis ini mengembalikan satu kolom dan satu baris sehingga hanya menghasilkan satu baris saja/nilai tunggal.

b. *Row subquery*

Jenis ini mengembalikan satu baris dengan banyak kolom. sehingga hanya menghasilkan satu baris saja/nilai tunggal. Terdapat operator khusus yang dapat membantu subquery jenis ini selain menggunakan bentuk standar subquery, operator tersebut adalah ANY, EXIST, IN, SOME, dan ALL.

c. *Nested subquery*

Jenis ini merupakan subquery yang berada didalam subquery lainnya dan tertanam pada klausa WHERE.

2.7 Natural Language Processing (NLP)

Natural Language Processing atau NLP merupakan penerapan di bidang ilmu komputer untuk mengkaji interaksi antara komputer dengan bahasa alami (bahasa manusia). Bahasa alami ini akan di proses dengan segala aturan gramatika dan

semantiknya dan mengubah bahasa tersebut sehingga dapat diproses oleh komputer[9].

Adapun tantangan yang harus dihadapi dalam penerapan NLP ini, yaitu:

a. Ambiguitas sintaksis

Ambiguitas semantik menjadi tantangan dalam penerapan NLP karena suatu bahasa memiliki berbagai kemungkinan struktur kalimat. Untuk memilih struktur yang paling tepat membutuhkan gabungan dari informasi semantik dan kontekstual.

b. Disambiguitas makna kata

Dalam bahasa alami terdapat kata-kata yang memiliki makna ganda / lebih dari satu makna seperti “bisa” dalam makna “dapat” dan dalam makna “racun”.

2.8 *Preprocessing*

Tahapan *preprocessing* ini merupakan tahapan awal untuk mengelola dan mempersiapkan kalimat masukan sebelum dilakukannya proses translasi. *Preprocessing* ini dilakukan untuk membersihkan kalimat yang dihasilkan bersih dari *noise* dan tidak berpengaruh untuk proses selanjutnya[10].

2.9 *Case Folding*

Proses *case folding* merupakan proses untuk menyeragamkan atau mengubah semua huruf dalam suatu kalimat/kata menjadi huruf kecil atau huruf kapital[11]. Dalam penelitian ini akan menggunakan *case folding* mengubah semua huruf pada menjadi huruf kecil.

Tabel 2.1 Case Folding

Sebelum	Sesudah
<u>T</u> ampilkan data <u>S</u> iswa yang berumur 13 tahun dan berasal dari <u>K</u> uningan!	<u>T</u> ampilkan data <u>s</u> iswa yang berumur 13 tahun dan berasal dari <u>k</u> uningan!

Pada tabel diatas kata ‘Tampilkan’, ‘Siswa’, dan ‘Bandung’ memiliki huruf kapital. Huruf kapital yang terdeteksi tersebut akan diubah menjadi huruf kecil sehingga menjadi ‘tampilkan’, ‘siswa’, dan ‘bandung’.

2.10 Filtering

Proses *Filtering* merupakan proses untuk mengurangi karakter yang tidak diperlukan pada kalimat masukan dan karakter yang menjadi masukan[12] hanya ‘a’ sampai ‘z’, ‘0’ sampai ‘9’. Tujuan dari proses filtering ini adalah untuk mengurangi karakter yang yang tidak dibutuhkan dan tidak mengganggu pada proses selanjutnya. Diharapkan juga dengan proses ini akna mengurangi noise.

Tabel 2.2 Filtering

Sebelum	Sesudah
tampilkan data siswa yang berumur 13 tahun dan berasal dari kuningan!	Tampilkan data siswa yang berumur 13 tahun dan berasal dari kuningan

Pada tabel diatas terdapat tanda seru (!) yang termasuk dalam kategori karakter masukan sehingga karakter tersebut akan di hapuskan.

2.11 Tokenizing

Tahap *tokenizing* merupakan tahapan untuk memisahkan kalimat menjadi token-token bisa berupa kata-kata, simbol atau angka[13]. Pada penelitian ini parameter pemisahannya berupa spasi untuk *Tokenizing* kata dan tanda titik untuk *Tokenizing* kalimat. Proses ini dibutuhkan untuk mendukung proses selanjutnya yg membutuhkan data berupa token-token. Hasil dari *Tokenizing* yaitu berupa token-token.

Tabel 2.3 Tokenizing

Sebelum	Sesudah
tampilkan data siswa yang berumur 13 tahun dan berasal dari kuningan	<ol style="list-style-type: none"> 1. tampilan 2. data 3. siswa 4. yang

	<ol style="list-style-type: none"> 5. berumur 6. 13 7. tahun 8. dan 9. berasal 10. dari 11. kuningan
--	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Pada tabel diatas proses dilakukan dengan mengecek karakter spasi untuk memisahkan kalimat menjadi token-token.

2.12 Stemming

Stemming adalah proses mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (*root word*) dengan menggunakan aturan-aturan tertentu[14] agar menghasilkan kata dasar. Proses ini dibutuhkan karena kata yang akan digunakan adalah kata dasar sehingga diperlukan proses penghapusan imbuhan. Contoh imbuhan yang akan dihilangkan ‘-kan’, ‘ber-‘ dan imbuhan lainnya.

Tabel 2.4 Stemming

Sebelum	Sesudah
1. tampilkan	1. tampil
2. data	2. data
3. siswa	3. siswa
4. yang	4. yang
5. berumur	5. umur
6. 13	6. 13
7. tahun	7. tahun
8. dan	8. dan
9. berasal	9. asal
10. dari	10. dari
11. kuningan	11. kuningan

Pada tabel diatas terdapat kata yang memiliki imbuhan ‘-kan’ yaitu pada kata ‘tampilkan’ akan di ubah menjadi ‘tampil’, imbuhan ‘ber-‘ pada kata ‘berumur’ dan ‘berasal’ akan diubah menjadi ‘umur’ dan ‘asal’. Sehingga kata-kata yg berimbuhan akan menjadi kata-kata dasar.

2.13 Stopword Removal

Pada tahap *stopword removal* ini merupakan proses untuk menghilangkan kata yang dianggap tidak penting untuk membangun struktur kalimat tanpa mewakili konten dari isi kalimat tersebut[15].

Tabel 2.5 *stopword removal*

Sebelum	Sesudah
1. tampilan	1. tampil
2. data	2. siswa
3. siswa	3. umur
4. yang	4. 13
5. berumur	5. dan
6. 13	6. asal
7. tahun	7. kuningan
8. dan	
9. berasal	
10. dari	
11. kuningan	

2.14 HTML

HTML merupakan salah satu format yang digunakan dalam pembuatan halaman web. Format dokumen dan gabungan dari hypertext yang menghubungkan dengan dokumen lain dapat disimpan pada komputer yang sama maupun komputer yang berbeda. HTML juga dapat dijalankan pada sistem operasi apapun. Menurut Williams dan Sawyer, Hypertext Markup Language (HTML) adalah sekumpulan perintah khusus (disebut “tag” atau “markup”) yang dipakai untuk menentukan struktur, bentuk dan link pada dokumen ke dokumen multimedia lain di website[16].

2.15 PHP

PHP merupakan salahsatu bahasa pemrograman yang digunakna untuk pembuatan dan pembangunan subuah web dan dapan digunakan bersama dengan skrip HTML. Bahasa PHP ini dapat disebut sebagai gabungan dari bahasa-bahasa pemrograman seperti *Purl*, *C*, dan *Java*. Adapun yang berpendapat bahwa merupakna *Hypertext Procesing* suatu sebuah bahasa pemrogramna yang tersiri dari kode-kode untuk pengolahan data dan mengirimkannya kembali menjadi kode HTML ke *web browser*[17].

Bahasa PHP ini dapat digunakan diberbagai sistem operasi dan mampu berkomunikasi dengan database. Dalam PHP kita dapat melakukan proses pengolaan database dengan penambahan, pengeditan, pencarian dan penghapusan.

2.16 XAMPP

XAMPP ini server yang berdiri sendiri, dapat digunakan di banyak sistem operasi dan merupakan kumpulan dari beberapa program. Program yang terdapat pada XAMPP yaitu Apache, HTTP server, MySQL, serta bahasa pemrograman PHP dan Perl sebagai penerjemah bahasa. Kegunaan aplikasi ini membuat jaringan lokal (website secara offline) menjadikan komputer kita sebagai server.

2.17 Pemodelan

Skema pemodelan data adalah suatu metode yang dapat memungkinkan untuk dijadikan model pada aplikasi *Data Manipulation Language* dari rancangan bangun penerjemah teks bahasa Indonesia.

2.17.1 Flowchart

Flowchart adalah bagan bagan yang mempunyai arus yang menggambarkan langkah – langkah penyelesaian sebuah masalah. Flowchart merupakan suatu cara dalam menyajikan suatu algoritma. Flowchart adalah suatu penggambaran secara grafik dari langkah – langkah dan urutan – urutan prosedur dari sebuah program.

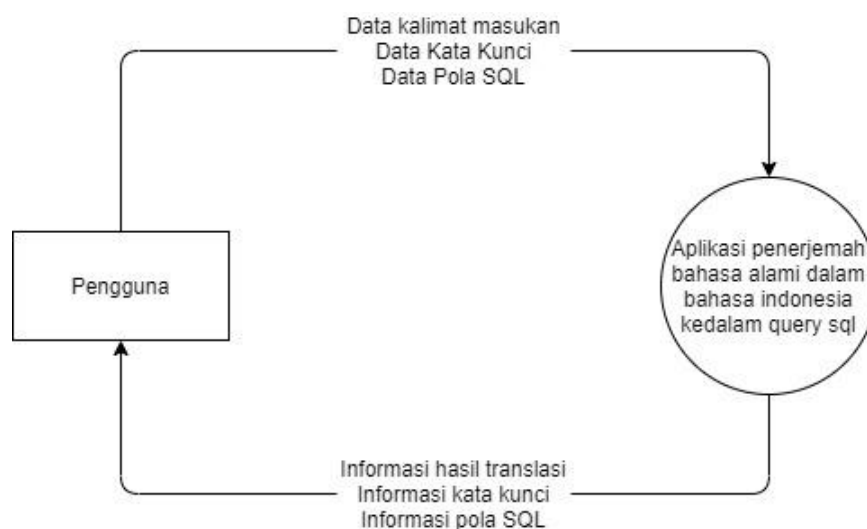
Flowchart biasanya digunakan untuk mempermudah menyelesaikan suatu masalah yang perlu dipelajari dan dievaluasi lebih lanjut khususnya.

Tujuan pembuatan flowchart adalah menggambarkan suatu tahapan penyelesaian masalah sederhana, terurai dan jelas menggunakan tombol – tombol standar.

2.17.2 Diagram Konteks

Diagram konteks merupakan sebuah diagram sederhana yang memberikan gambaran hubungan antara entity luar, keluaran dan masukan dari sistem. Diagram konteks diwujudkan dengan lingkaran tunggal yang mewakili seluruh sistem. Diagram konteks dimulai dengan menggambarkan terminator, aliran data, aliran kontrol penyimpana, dan proses tunggal yang menunjukkan seluruh sistem. Terdapat beberapa aturan dalam penggambaran diagram konteks. Berikut merupakan aturan – aturan dalam penggambaran diagram konteks :

1. Hanya menggunakan satu simbol proses
2. Untuk menggambarkan seluruh sistem diberi label proses yang berupa kata kerja ditambah objek
3. Tidak memberikan nomor pada simbol proses
4. Mengikutsertakan semua terminator dari sistem
5. Menunjukkan seluruh arus data antara terminator dan sistem



Gambar 2.1 Diagram Konteks

2.17.3 DFD (*Data Flow Diagram*)

DFD adalah suatu model logika data atau proses yang dibuat untuk penggambaran darimana asal data kemana tujuan data yang keluar dari sistem, data disimpan dimana, proses apa yang menghasilkan data tersebut dan interaksi antara data yang tersimpan dan proses yang dikenakan pada data.

1. Arus Data (*Data Flow*)

Menunjukkan arus dari data yang merupakan masukan untuk sistem atau proses sistem.

2. Proses

Proses merupakan suatu kegiatan yang dilakukan oleh orang, komputer dari hasil dari arus data yang masuk kedalam proses untuk dihasilkan arus data yang akan keluar dari proses.

3. Kesatuan Luar (*External Entity*)

Kesatuan luar adalah kesatuan di lingkungan diluar sistem yang dapat berupa orang, sistem lain, ataupun organisasi yang akan memberikan masukan atau menerima keluaran dari sistem.

4. File

Kumpulan data yang disimpan dengan cara tertentu. Data yang mengalir disimpan dalam file. Aliran data ditambahkan kedalam file. Pada penelitian kali ini DFD atau *Data Flow Diagram* digunakan untuk menggambarkan rancangan dan alur proses sistem secara detail yang dapat dengan mudah dipahami oleh pengguna maupun yang membuat sistem.

2.17.4 Kamus Data

Kamus data merupakan kumpulan daftar elemen data yang mengalir pada sistem perangkat lunak sehingga masukan (input) dan keluaran (output) dapat dimengerti secara umum. Kamus data mendefinisikan elemen data dengan menjelaskan arti aliran data dan penyimpanan data dalam DFD, mendeskripsikan komposisi paket data yang bergerak melalui aliran, mendeskripsikan komposisi

penyimpanan data, menspesifikasikan nilai dan satuan yang relevan bagi penyimpanan dan aliran, mendeskripsikan hubungan detil antar penyimpanan.

Kamus data berfungsi untuk membantu pelaku sistem untuk mengartikan alokasi secara detail dan mengorganisasikan semua elemen data yang digunakan dalam sistem secara persis sehingga baik pemakai atau penganalisis sistem mempunyai dasar pengertian yang sama tentang masukan, keluaran, penyimpanan dan proses. Kamus data biasanya berisi:

1. Nama – nama dari data
2. Deskripsi yang merupakan deskripsi data
3. Digunakan pada – merupakan proses-proses yang terkait data

Informasi tambahan – seperti tipe data, nilai data, batas nilai data, dan komponen yang membentuk data.

2.17.5 Pengujian

Untuk dapat menghitung akurasi dari hasil query dilakukan perhitungan dengan rumus berikut.

$$\text{Akurasi} = \frac{\text{Jumlah hasil query sistem benar}}{\text{Jumlah sampel data uji}} \times 100\%$$

Pengujian pada penelitian ini memiliki tujuan untuk menghitung akurasi yang digunakan untuk mengetahui tingkat keberhasilan dari proses translasi yang dilakukan oleh sistem dengan membandingkan query harapan dan query hasil translasi.