

BAB 2

TINJAUAN PUSTAKA

2.1. Landasan Teori

Landasan teori merupakan definisi, dan konsep yang berkaitan dengan pembangunan Konversi Bahasa Indonesia Ke Bahasa Query Terstruktur. Landasan teori ini akan menjadi dasar yang kuat dalam penelitian yang akan dilakukan.

2.2. Aplikasi

Aplikasi dapat diartikan sebagai suatu program berbentuk perangkat lunak yang berjalan pada suatu sistem tertentu yang berguna untuk membantu berbagai kegiatan yang dilakukan oleh manusia. Selain pengertian tersebut ada banyak pengertian dari para ahli mengenai kata aplikasi adalah sebagai berikut[10]:

a. Ali Zaki dan Smitdev Community

Menurut Ali Zaki dan Smitdev Community, Aplikasi merupakan komponen yang bermanfaat sebagai media untuk menjalankan pengolahan data ataupun berbagai kegiatan lainnya seperti pembuatan ataupun pengolahan dokumen dan file.

b. Sri Widianti

Menurut Sri Widianti, Aplikasi merupakan sebuah software (perangkat lunak) yang bertugas sebagai front end pada sebuah sistem yang dipakai untuk mengelola berbagai macam data sehingga menjadi sebuah informasi yang bermanfaat untuk penggunanya dan juga sistem yang berkaitan.

c. Harip Santoso

Menurut Harip Santoso, Aplikasi merupakan sebuah kelompok file (class, form, report) yang ditujukan sebagai pengeksekusi aktivitas tertentu yang saling berkaitan seperti contohnya aplikasi payroll dan aplikasi fixed asset.

d. Yuhefizer

Menurut Yuhefizar, Aplikasi adalah program yang sengaja dibuat dan dikembangkan sebagai pemenuh kebutuhan penggunanya dalam menjalankan suatu pekerjaan tertentu.

e. Hengky W.Pranama

Menurut Hengky W. Pramana, pengertian aplikasi adalah satu unit perangkat lunak yang sengaja dibuat untuk memenuhi kebutuhan akan berbagai aktivitas ataupun pekerjaan, seperti aktivitas perniagaan, periklanan, pelayanan masyarakat, game, dan berbagai aktivitas lainnya yang dilakukan oleh manusia.

Dengan berkembang pesatnya teknologi saat ini, media informasi yang beredar sebagian besar menggunakan aplikasi. Maka dari itu penelitian ini nantinya akan menghasilkan sebuah aplikasi.

2.3. HTML

Hypertext Markup Language (HTML) adalah sebuah bahasa markah yang digunakan untuk membuat sebuah halaman web, menampilkan berbagai informasi di dalam sebuah penjelajah web Internet dan pemformatan hiperteks sederhana yang ditulis dalam berkas format ASCII agar dapat menghasilkan tampilan wujud yang terintegrasi. Dengan kata lain, berkas yang dibuat dalam perangkat lunak pengolah kata dan disimpan dalam format ASCII normal sehingga menjadi halaman web dengan perintah-perintah HTML. 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)[11].

2.4. PHP

PHP adalah singkatan dari "PHP: Hypertext Preprocessor", yaitu Bahasa script open source yang sederhana namun kuat untuk menciptakan konten web dinamis. Jutaan situs web yang didukung oleh PHP adalah bukti popularitas dan kemudahan penggunaannya. PHP digunakan oleh kedua pemrogram, yang menghargai fleksibilitas dan kecepatan, dan perancang web, yang menghargai aksesibilitas dan kenyamanannya. PHP diciptakan oleh Rasmus Lerdorf pertama kali tahun 1994. Pada awalnya PHP adalah singkatan dari "Personal Home Page Tools". Selanjutnya diganti menjadi FI ("Forms Interpreter"). Sejak versi 3.0, nama bahasa ini diubah menjadi "PHP: Hypertext Preprocessor" dengan singkatannya "PHP". PHP versi terbaru adalah versi ke-5. Berdasarkan survey Netcraft pada bulan Desember 1999, lebih dari sejuta website menggunakan PHP, di antaranya adalah NASA, Mitsubishi, dan RedHat. [12].

2.5. SQL

Structure Query Language atau SQL merupakan suatu bahasa yang digunakan untuk mengakses database. SQL juga sering disebut query. Secara umum SQL terdiri dari dua bahasa, yaitu Data Definition Language (DDL) dan Data Manipulation Language (DML). DDL digunakan untuk mendefinisikan, mengubah, serta menghapus database dan objek – objek yang diperlukan database. Secara umum, DDL yang digunakan adalah Create, Alter, Use, Drop yang biasanya digunakan oleh administrator dalam pembuatan suatu aplikasi database. DML digunakan untuk memanipulasi data yang ada dalam suatu tabel. Perintah yang dapat dilakukan adalah Select, Insert, Update, Delete. Fungsi Agregat, yaitu fungsi khusus yang melibatkan sekelompok data, secara umum fungsi agregat adalah SUM, COUNT, AVG, MAX, dan MIN. Fungsi ini digunakan pada bagian select, syarat untuk fungsi ini adalah diletakkan pada bagian having, bukan where. Sub query adalah suatu query yang sudah menjadi kompleks terutama melibatkan lebih dari satu tabel atau fungsi agregat [13].

2.6. NLP

Pemrosesan bahasa alami (Natural Language Processing) adalah salah satu bidang ilmu kecerdasan buatan (Artificial Intelligence) yang mengkaji tentang komunikasi antar manusia dengan komputer melalui bahasa alami baik secara lisan maupun tulisan.

Proses representasi bahasa dalam bentuk yang lebih memungkinkan untuk dikomputasi telah dilakukan oleh seorang ahli bernama Chomsky pada tahun 1957. Pada bidang ilmu pemrosesan bahasa alami tidak mudah untuk dilakukan karena terdapat beberapa alasan yang menyulitkan pada pemrosesan bahasa alami. Salah satu kesulitan dalam pemrosesan bahasa alami adalah pemilihan arti yang tepat dari suatu kata bermakna ganda seperti kata 'bisa' yang dapat berarti 'racun' atau 'dapat' sesuai dengan kalimatnya. Fenomena ini terjadi dalam penentuan jenis kata (part of speech) seperti kata 'advanced' yang dapat berfungsi sebagai kata kerja aktif lampau, kata kerja pasif, atau kata sifat. Selain itu, jumlah kosa kata (vocabulary) dalam bahasa alami besar dan terus berkembang dari waktu ke waktu.

Pemrosesan Bahasa Alami sering diterapkan pada aplikasi tertulis yang sering menemui kesulitan karena tulisan setiap orang tidak selalu baku sesuai dengan tata bahasa sehingga tidak sesuai dengan aturan – aturan yang berlaku. Salah satu aplikasi dari pemrosesan bahasa alami adalah sistem penerjemah bahasa alami (natural language translator) seperti sistem penerjemah dari bahasa Inggris ke bahasa Indonesia. Translator tidak hanya dapat menerjemahkan kata per kata (kamus) tetapi juga dapat mentranslasikan dari bahasa asal ke bahasa target dengan maksud yang dapat dimengerti [14].

2.7. Adobe Dreamweaver

Adobe Dreamweaver adalah aplikasi desain dan pengembangan web yang menyediakan editor WYSIWYG visual (bahasa sehari-hari yang disebut sebagai Design view) dan kode editor dengan fitur standar seperti syntax highlighting, code completion, dan code collapsing serta fitur lebih canggih seperti real-time syntax checking dan code introspection untuk menghasilkan petunjuk kode untuk membantu pengguna dalam menulis kode. Tata letak tampilan Design memfasilitasi

desain cepat dan pembuatan kode seperti memungkinkan pengguna dengan cepat membuat tata letak dan manipulasi elemen HTML. Dreamweaver memiliki fitur browser yang terintegrasi untuk melihat halaman web yang dikembangkan di jendela pratinjau program sendiri agar konten memungkinkan untuk terbuka di web browser yang telah terinstall. Aplikasi ini menyediakan transfer dan fitur sinkronisasi, kemampuan untuk mencari dan mengganti baris teks atau kode untuk mencari kata atau kalimat biasa di seluruh situs, dan templating feature yang memungkinkan untuk berbagi satu sumber kode atau memperbarui tata letak di seluruh situs tanpa server side includes atau scripting. Behavior Panel juga memungkinkan penggunaan JavaScript dasar tanpa pengetahuan coding, dan integrasi dengan Adobe Spry Ajax framework menawarkan akses mudah ke konten yang dibuat secara dinamis dan interface.

Dreamweaver dapat menggunakan ekstensi dari pihak ketiga untuk memperpanjang fungsionalitas inti dari aplikasi, yang setiap pengembang web bisa menulis (sebagian besar dalam HTML dan JavaScript).

Pada versi 5, Dreamweaver mendukung syntax highlighting untuk bahasa seperti berikut:

1. ActionScript
2. Active Server Pages (ASP).
3. C#
4. Cascading Style Sheets (CSS)
5. ColdFusion
6. EDML
7. Extensible HyperText Markup Language (XHTML)
8. Extensible Markup Language (XML)
9. Extensible Stylesheet Language Transformations (XSLT)
10. HyperText Markup Language (HTML)
11. Java
12. JavaScript
13. PHP: Hypertext Preprocessor (PHP)

14. Visual Basic (VB)
15. Visual Basic Script Edition (VBScript)
16. Wireless Markup Language (WML)

2.8. DFD

DFD adalah suatu model logika data atau proses yang dibuat untuk menggambarkan darimana asal data dan kemana tujuan data yang keluar dari sistem, dimana data disimpan, proses apa yang menghasilkan data tersebut dan interaksi antara data yang tersimpan dan proses yang dikenakan pada data tersebut [15].

1. Arus Data (Data Flow)

Menunjukkan arus dari data yang dapat berupa masukan untuk sistem atau dari proses sistem.

2. Proses

Proses adalah kegiatan yang dilakukan oleh orang, komputer dari hasil arus data yang masuk ke dalam proses untuk dihasilkan arus data yang akan keluar dari proses.

3. Kesatuan Luar (External Entity)

Kesatuan luar merupakan kesatuan di lingkungan luar sistem yang dapat berupa orang, organisasi atau sistem lain 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 di-update atau ditambahkan ke dalam file.

Pada penelitian ini Data Flow Diagram (DFD) digunakan untuk menggambarkan rancangan dan alur proses sistem secara detail yang mudah dipahami oleh pemakai maupun pembuat sistem.

2.9. Kalimat

Kalimat adalah satuan bahasa yang disusun konstituen dasar yang umumnya berupa klausa, kata penghubung (jika ada) dan intonasi final. Dalam ragam tulisan

intonasi final ini dinyatakan dengan titik (.), tanda tanya (?), atau tanda seru (!) [16]. Menurut strukturnya, sebuah kalimat sederhana dalam bahasa Indonesia memiliki pola sebagai berikut [17].

Tabel 2. 1 Pola Kalimat Sederhana Berdasarkan Strukturnya

No	Struktur	Contoh
1	Subjek + Predikat (S-P)	1. Para pedagang mengeluh. 2. Orang itu sedang tidur.
2	Subjek + Predikat + Objek (S-P-O)	1. Ilmu mengatur pengetahuan. 2. Ayahnya membeli mobil baru.
3	Subjek + Predikat + Pelengkap (S-P-Pel)	Beliau menjadi ketua koperasi.
4	Subjek + Predikat + Objek + Pelengkap (S-P-O-Pel)	1. Ia menyebutnya si hitam. 2. Dian mengambilkan adiknya air minum.
5	Subjek + Predikat + Objek + Keterangan (S-P-O-K)	Pak Raden memasukan uang ke bank.

6	Subjek + Predikat + Keterangan (S-P-K)	Kami tinggal di jakarta
---	---	----------------------------

Dalam tata bahasa Indonesia, fungsi gramatikal subjek, predikat, dan objek dapat ditempati oleh berbagai macam jenis kata maupun frasa. Tabel 2.2 berikut adalah daftar fungsi gramatikal dan jenis kata atau frasa yang dapat menempatinnya.

Tabel 2.2 Daftar Fungsi Gramatikal dan Pengisiannya

Fungsi Gramatikal	Keterangan	Pengisi
Subjek	Sebuah kata atau frasa yang letaknya di depan predikat.	<ol style="list-style-type: none"> 1. Kata benda / nomina 2. Frasa nomina 3. Kata kerja / verba 4. Kata bilangan / numeralia
Predikat	Bagian dari klausa yang memberikan informasi tentang subjek.	<ol style="list-style-type: none"> 1. Kata Kerja / verba 2. Frasa verba 3. Adjektiva 4. Kata benda / nomina 5. Frasa nomina 6. Kata / frasa numeralia
Objek	Bagian dari klausa yang terkena tindakan yang disebut di dalam predikat.	<ol style="list-style-type: none"> 1. Kata benda / nomina 2. Frasa nomina
Pelengkap	Bagian dari klausa	<ol style="list-style-type: none"> 1. Adverbialia

	yang mengikuti atau melengkapi predikat.	2. Frasa adverbial
Keterangan	Bagian dari klausa yang menjelaskan jenis tindakan yang dinyatakan di dalam predikat. Biasanya keterangan berupa kata keterangan. Posisinya dalam kalimat bisa diubah (depan,tengah, belakang).	Frasa nomina

Berdasarkan jumlah subjek dan predikatnya (jumlah klausa), kalimat dikelompokkan menjadi kalimat tunggal dan kalimat majemuk [17].

2.10. Kalimat Tunggal

Kalimat tunggal adalah kalimat yang terdiri atas satu klausa. Kalimat ini hanya mempunyai satu subjek dan satu predikat. Kalimat tunggal bisa berwujud kalimat dasar atau kalimat dasar yang diperluas, meskipun sudah mengalami perluasan, fungsi subjek dan predikatnya tetap berjumlah satu [17].

Contoh :

- a. Separuh pesisir Pulau Bangka Rusak.
- b. Separuh pesisir Pulau Bangka, Provinsi Kepulauan Bangka Belitung, rusak.

Kalimat pada contoh (a) merupakan kalimat dasar, sedangkan kalimat pada contoh (b) merupakan kalimat tunggal hasil perluasan karena subjek kalimat telah diperluas oleh unsur keterangan. Meskipun demikian, subjek pada kalimat tersebut

tetap berjumlah satu.

2.11. Kalimat Majemuk

Kalimat majemuk adalah kalimat yang terdiri atas dua klausa atau lebih yang dimana antara klausa yang satu dengan klausa yang lain saling berhubungan. Umumnya, ahli bahasa membagi hubungan antar klausa ke dalam dua jenis yaitu koordinasi dan subordinasi.

1. Hubungan koordinasi (Majemuk Setara)

Hubungan koordinasi menggabungkan dua klausa atau lebih yang masing – masing mempunyai kedudukan setara dalam struktur kalimat. Kalimat yang terbentuk dari dua klausa atau lebih dengan hubungan antar klausa terbentuk secara koordinatif dapat disebut dengan kalimat majemuk setara, sehingga kalimat tersebut terdiri dari klausa – klausa yang disusun secara koordinasi dengan kata penghubung yang dimana setiap klausa dapat berdiri sendiri menjadi kalimat.

Contoh :

“Candi Gedong Songo memiliki sembilan kelompok candi, tetapi sebagian kelompok candi itu sudah hilang.”

Kalimat diatas terdiri atas dua klausa : (a) Candi Gedong Songo memiliki sembilan kelompok candi; (b) sebagian kelompok candi itu sudah hilang. Kalimat tersebut dihubungkan oleh secara koordinasi oleh kata ‘tetapi’ sehingga terbentuk kalimat majemuk setara.

2. Hubungan Subordinasi (Majemuk Bertingkat)

Hubungan subordinasi menunjukkan hubungan yang hierarkis, yakni menggabungkan dua klausa atau lebih secara bertingkat, ada yang berfungsi sebagai klausa utama dan ada yang berfungsi sebagai klausa bawahan. Kalimat yang terbentuk dari beberapa klausa yang dihubungkan dengan kata penghubung yang bersifat subordinatif menyebabkan klausa yang satu menjadi bagian dari klausa yang lain, artinya salah satu klausa bergantung pada klausa lain dan tidak bisa berdiri sendiri. Kalimat tersebut disebut dengan kalimat majemuk bertingkat.

Contoh :

“Candi Gedong Songo merupakan mutiara kehidupan karena menjadi sumber nafkah bagi masyarakat sekitarnya.”

Kalimat di atas terdiri dari dua klausa, yaitu (a) “Candi Gedong Songo merupakan mutiara kehidupan”, dan (b) “(Candi Gedong Songo) menjadi sumber nafkah bagi masyarakat sekitarnya”. Kedua klausa dihubungkan oleh konjungsi ‘karena’. Klausa (a) merupakan klausa utama, sedangkan klausa (b) merupakan klausa bawahan [17].

Daftar kata penghubung dan ciri – cirinya dapat dilihat pada Tabel 2.3.

Tabel 2.3 Daftar Kata Penghubung Kalimat Majemuk

No	Kalimat Majemuk	Penghubung Klausa	Ciri-Ciri Penghubung
1	Majemuk setara / Koordinatif	1. dan 2. atau 3. tapi 4. namun 5. serta 6. lalu 7. kemudian 8. bahkan	lokasi penghubung tetap, yaitu di awal klausa koordinatif.
2	Majemuk bertingkat / Subordinatif	1. karena / sebab 2. ketika 3. sejak 4. jikalau 5. asalkan 6. meskipun / walaupun 7. bila / apabila 8. maka 9. sehingga	Subordinator selalu mengikuti unsur intern klausa utama sebagai informasi lanjutan. Karena fungsinya itu, subordinator selalui terletak di awal klausa subordinatif / bawahan.

2.12. Kalimat Perintah

Kalimat perintah adalah kalimat yang isinya mengharapkan adanya reaksi berupa tindakan atau perbuatan dari orang yang diajak bicara (pendengar atau pembaca) [18]. Kalimat perintah dilihat dari taraf tindakan yang diharapkan dapat dibedakan menjadi 3 yaitu kalimat perintah yang tegas, kalimat perintah yang biasa, dan kalimat perintah yang halus.

a. Kalimat perintah tegas

Kalimat perintah yang tegas dibentuk dari sebuah klausa tidak lengkap, biasanya hanya berupa kata kerja dasar, disertai dengan intonasi kalimat perintah dan dalam bahasa tulis, intonasi kalimat perintah ini diganti atau dilambangkan dengan tanda seru. Contoh :

1. Tangkap!
2. Tembak!

Jika contoh dilengkapi dengan objek atau keterangan, agar lebih jelas atau agar tidak menimbulkan salah paham. Misalnya contoh yang sebelumnya dibuat menjadi “Tangkap orang itu!”.

b. Kalimat perintah biasa

Kalimat perintah yang biasa dibentuk dari sebuah klausa berpredikat kata kerja dasar yang diberi partikel “-lah”, serta dengan menghilangkan atau melesapkan subjeknya. Contoh :

1. Jagalah kebersihan!
2. Bayarlah dengan uang pas!
3. Belilah karcis di loket!

Jika kalimat perintah digunakan pada orang yang sudah ditentukan, maka subjek pada klausa tersebut harus ditampilkan, misalnya contoh yang sebelumnya dibuat menjadi “Beni, jagalah kebersihan!”.

c. Kalimat perintah halus

Kalimat perintah yang halus harus digunakan oleh yang lebih muda kepada yang lebih tua, yang lebih rendah status atau kedudukan sosialnya terhadap

yang lebih tinggi atau yang lebih berkuasa, atau juga untuk menampilkan rasa hormat atau sopan santun terhadap orang yang diperintah. Contoh :

1. Tolong bereskan buku – buku ini!
2. Silahkan duduk di sini!

2.13. Case Folding

Case folding adalah proses penyeragaman semua teks ke dalam case yang sama menjadi huruf kecil (lowercase) atau huruf capital (uppercase) [19]. Pada penelitian ini setiap huruf dalam teks masukan diseragamkan menjadi huruf kecil. Contoh dari tahap case folding dapat dilihat pada tabel 2.4. Tabel tersebut menjelaskan bahwa setiap huruf kapital yang ada pada teks input akan diubah menjadi huruf kecil.

Tabel 2. 4 Contoh Tahap Case Folding

Kalimat Masukan	Sesudah Case Folding
Buatkan Database Dengan Nama Mahasiswa	<u>b</u> uatkan <u>d</u> atabase <u>d</u> engan <u>n</u> ama <u>m</u> ahasiswa

Penelitian ini menggunakan tahap – tahap proses case folding untuk menyeragamkan semua huruf pada kalimat masukan menjadi huruf kecil (lowercase).

2.14. Filtering

Filtering merupakan proses pengecekan pada setiap karakter pada teks selain karakter ‘a’ sampai ‘z’, ‘0’ sampai ‘9’, ‘_’, ‘,’ dan spasi akan dihapus. Berikut adalah contoh pada tahap filtering.

Tabel 2.5 Contoh Tahap Filtering

Kalimat Masukan	Sesudah filtering
buatkan ‘_database_’ dengan nama mahasiswa!	buatkan database dengan nama dbmahasiswa

Penelitian ini menggunakan tahap filtering untuk memindai setiap karakter lalu menghapus karakter lain selain karakter ‘a’ sampai ‘z’, ‘0’ sampai ‘9’, ‘_’, ‘,’ dan spasi.

2.15. Tokenizing Kata

Tokenizing kata adalah proses memindai teks masukan yang akan dipotong menjadi kumpulan kata. Pemecahan kalimat menjadi kata – kata tunggal dilakukan dengan memindai kalimat dan memisahkan kata demi kata dengan spasi. Hasil keluaran dari proses tokenizing kata akan digunakan sebagai masukan dalam tahap transformasi teks. Contoh tokenizing kata dapat dilihat pada tabel 2.6.

Tabel 2.6 Contoh Tahap Tokenizing Kata

Kalimat Masukan	Sesudah Tokenizing Kata				
buatkan database dengan nama mahasiswa	buatkan	database	dengan	nama	mahasiswa

Penelitian ini menggunakan tahap tokenizing kata untuk memotong teks masukan menjadi kumpulan kata.

2.16. Stemming

Stemming merupakan suatu proses yang mentransformasi kata-kata yang berimbuhan ke kata-kata dasarnya (root word) [19]. Sebagai contoh, kata ‘menyenangkan’, ‘menyenangi’, dapat dikatakan serupa atau satu kelompok dan dapat diwakili oleh satu kata umum ‘senang’. Proses stemming pada teks berbahasa Indonesia berbeda dengan stemming pada teks berbahasa Inggris. Pada teks berbahasa Inggris, proses yang diperlukan hanya proses menghilangkan sufiks, sedangkan pada teks berbahasa Indonesia, selain sufiks, prefiks, dan konfiks juga dihilangkan. Pada umumnya kata dasar pada bahasa Indonesia terdiri dari kombinasi:

$$\text{Prefiks 1} + \text{Prefiks 2} + \text{Kata dasar} + \text{Sufiks 3} + \text{Sufiks 2} + \text{Sufiks 1} \quad (2.1)$$

Pada penelitian ini, algoritma yang digunakan dalam proses stemming yaitu Nazief dan Andriani yang telah dikembangkan oleh Sastrawi. Algoritma Stemming Bahasa Indonesia M. Adriani dan B. Nazief ini mempunyai aturan imbuhan sendiri dengan model sebagai berikut.

$$[[[AW+]AW+]AW+] \text{ Kata-Dasar } [[+AK][+KK][+P]] \quad (2.2)$$

Dimana:

AW : Awalan

AK : Akhiran

KK : Kata Ganti kepunyaan

P : Partikel

Algoritma yang dibuat oleh Bobby Nazief dan Mirna Adriani ini memiliki tahap sebagai berikut.

1. Mencari kata yang akan diproses stemming dalam kamus. Jika ditemukan maka diasumsikan bahwa kata tersebut adalah root word, maka algoritma berhenti.
2. Inflection Suffixes (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa particles (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus Possesive Pronouns (“-ku”, “-mu”, atau “-nya”), jika ada.
3. Hapus Derivation Suffixes (“-i”, “-an” atau “-kan”). Jika kata ditemukan dikamus, maka algoritma berhenti. Jika tidak maka ke langkah 3a.
 - a. Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
 - b. Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
4. Hapus Derivation Prefix. Jika pada langkah 3 ada sufiks yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b.
 - a. Periksa tabel kombinasi awalan-akhirian yang tidak diijinkan. Jika ditemukan maka algoritma berhenti, jika tidak pergi ke langkah 4b.

b. For $i = 1$ to 3, tentukan tipe awalan kemudian hapus awalan. Jika root word belum juga ditemukan lakukan langkah 5, jika sudah maka algoritma berhenti. Catatan: jika awalan kedua sama dengan awalan pertama algoritma berhenti.

5. Melakukan Recoding.

6. Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai root word dan proses dianggap selesai.

Tipe awalan ditentukan melalui langkah-langkah sebagai berikut.

1. Jika awalannya adalah “di-”, “ke-”, atau “se-” maka tipe awalannya secara berturut-turut adalah “di-”, “ke-”, atau “se-”.

2. Jika awalannya adalah “te-”, “me-”, “be-”, atau “pe-” maka dibutuhkan sebuah proses tambahan untuk menentukan tipe awalannya.

3. Jika dua karakter pertama bukan “di-”, “ke-”, “se-”, “te-”, “be-”, “me-”, atau “pe-” maka berhenti.

4. Jika tipe awalan adalah “none” maka berhenti. Jika tipe awalan adalah bukan “none” maka awalan ditemukan, hapus awalan jika ditemukan.

Sebagai contoh dapat dilihat pada tabel 2.7 kata ‘buatkan’ akan diubah menjadi kata ‘buat’ karena pada kata ‘buatkan’ terdapat imbuhan “-kan” [20].

Tabel 2.7 Contoh Tahap Stemming

Kalimat Masukan	Sesudah Stemming
<u>buatkan</u> database dengan nama mahasiswa	<u>buat</u> database dengan nama mahasiswa

Pada penelitian ini digunakan library stemming dengan PHP untuk bahasa Indonesia dari Sastrawi. Karena library ini menggunakan tiga algoritma yaitu algoritma Nazief dan Adriani, algoritma CS Stemmer dan algoritma ECS Stemmer. Dimana algoritma CS Stemmer dan algoritma ECS Stemmer memodifikasi beberapa aturan yang terdapat di algoritma Nazief dan Adriani. Pada penelitian ini juga beberapa kata tidak dilakukan proses stemming karena proses stemming hanya untuk mendapatkan kata kunci.

2.17. Removing Stopword

Stopword merupakan kata yang diabaikan dalam pemrosesan, kata-kata ini biasanya disimpan ke dalam stop lists. Karakteristik utama dalam pemilihan stopwords biasanya adalah kata yang mempunyai frekuensi kemunculan yang tinggi misalnya kata penghubung seperti ‘dan’, ‘atau’, ‘tapi’, ‘akan’ dan lainnya. Tidak ada aturan pasti dalam menentukan stopwords yang akan digunakan, penentuan stopwords bisa disesuaikan dengan kasus yang sedang diselesaikan [20]. Beberapa kata yang termasuk stopwords dapat dilihat pada tabel 2.8.

Tabel 2.8 Beberapa Daftar Kata Yang Termasuk Stopword

No	Kata Stopword	No	Kata Stopword
1	coba	11	untuk
2	tolong	12	dengan
3	ada	13	apa
4	agar	14	maka
5	akan	15	lagi
6	berapa	16	lain
7	dan	17	jadi
8	yang	18	jika
9	pada	19	maka
10	yaitu	20	itu

Pada tabel 2.8 merupakan beberapa daftar kata yang termasuk stopwords, penelitian ini hanya menggunakan 320 kata dasar dari 758 kata stopwords yang digunakan oleh Fadillah Z Tala yang dapat dilihat pada lampiran B-1 dan disesuaikan dengan menghapus 3 kata (“buat”, “tambah”, dan “semua”) karena kata tersebut digunakan pada kamus kata kunci untuk menentukan perintah DDL dan menambah 3 kata (“tolong”, “coba”, “menjadi”) karena kata tersebut tidak digunakan pada kamus kata kunci.

