

PERANCANGAN TRANSLATOR BAHASA ALAMI KE DALAM FORMAT SQL (STRUCTURED QUERY LANGUAGE)

Kuspriyanto, Herry Sujani, Hary Tjahjono, Shofyan Kusuma

Departemen Teknik Elektro, Institut Teknologi Bandung

Jalan Ganesha No. 10 Bandung

E-mail: kuspriyanto@yahoo.com

ABSTRACT

In order to able to use SQL (Structured Query Language) properly, users should understand structure of SQL. In this research, a prototype of translator from natural language to SQL format is developed to facilitate novice users to get information from a database. Using three databases, testing results indicate that the translator gives accuracy of 86%.

Keywords: natural language translator, SQL

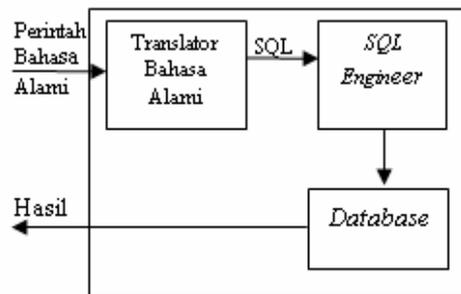
1. PENDAHULUAN

Basis data pada umumnya diaplikasikan dengan arsitektur Client-Server, dimana pada arsitektur ini terdapat dua komponen utama, yaitu *client* dan *server*. *Client* berisi aplikasi basis data sedangkan *server* berisi DBMS dan basis data. Setiap aktifitas yang dikehendaki pemakai akan lebih dahulu ditangani oleh *client*. Jika ada proses yang harus melibatkan data pada basis data, barulah *client* melakukan kontak permintaan terhadap *server* dan selanjutnya *server* mengeksekusi permintaan tersebut terhadap basis data yang terdapat di dalamnya [7].

Permintaan data oleh *client* terhadap *server* dinyatakan dengan bahasa basis data. Dari sekian banyak bahasa basis data yang pernah ada seperti SQL, dbase, QUEL dan lain-lain, yang menjadi standar pada saat ini adalah SQL (*Structured Query Language*).

Untuk melihat informasi dengan kriteria tertentu dalam suatu *database* dengan menggunakan SQL, maka pemakai harus memahami struktur SQL, dimana setelah diolah dalam "SQL engine", informasi pada *database* dapat dikeluarkan sebagai hasil *query*.

Untuk memudahkan pemakai mendapatkan informasi dalam suatu *database*, cukup menggunakan bahasa alami yang biasa digunakan sehari-hari. Agar hal ini dapat terjadi maka diperlukan suatu sub sistem lain diantara user dan sistem yang dapat mentranslasi bahasa alami ke dalam format SQL. Sub sistem yang dimaksud adalah "Translator bahasa alami ke SQL" seperti terlihat pada Gambar 1.



Gambar 1. Diagram blok *query* basisdata dengan bahasa alami

"Perintah bahasa alami" yang menjadi input sub sistem "Translator bahasa alami ke SQL" dalam sistem ini merupakan teks kalimat bahasa Indonesia yang baku, sedangkan "SQL" yang menjadi keluarannya berupa baris perintah dalam format SQL.

Proses aplikasi yang akan menerima keluaran sub sistem tersebut adalah proses pencarian data dengan format SQL. Proses pencarian data merupakan aplikasi SQL yang paling banyak digunakan dalam interaksi antara manusia dengan komputer, sedangkan aplikasi lainnya seperti penambahan, perubahan atau penghapusan data. biasanya hanya diijinkan untuk orang-orang tertentu.

Paper ini membahas tentang pengembangan metode pendektesian kata untuk melakukan proses translasi dari bahasa alami menjadi bahasa dalam format SQL.

2. SQL

Struktur dasar dari SQL terdiri dari tiga klausa yaitu *select*, *from*, *where*. Klausa *select* digunakan untuk menetapkan daftar atribut (*file*), yang diinginkan hasil *query*, klausa *from* digunakan untuk menetapkan tabel (*gabungan tabel*) yang akan ditelusuri selama *query* data dilakukan, sedangkan klausa *where* bersifat opsional, digunakan sebagai kriteria yang dipenuhi dalam memperoleh hasil *query*.

Sintaks (cara penulisan) dari struktur SQL dasar dengan tiga klausa tersebut adalah [2]:

```

Select select_list
from teble[ ,table...]
[where condition]
  
```

dimana

- *select_list* merupakan daftar atribut(*field*)
- *table* merupakan daftar tabel
- *condition* merupakan predikat *query*
- [] merupakan tanda optional (boleh digunakan atau tidak)

3. NLP (*NATURAL LANGUAGE PROCESSING*)

Secara umum NLP dapat didefinisikan sebagai kemampuan suatu komputer untuk memproses bahasa lisan atau tulisan yang digunakan oleh manusia dalam percakapan sehari-hari [5].

Bahasa merupakan fenomena yang rumit yang melibatkan proses sesukar pengenalan bunyi atau huruf, sintaksis kalimat, inferensi semantik tingkat tinggi, bahkan komunikasi emosi lewat irama bicara. Untuk mengelola kerumitan ini para ahli bahasa telah mendefinisikan tingkat analisis yang berbeda untuk bahasa alami yaitu persajakan, fonologi, morfologi, sintaks, semantik, pragmatik, dan pengetahuan tentang dunia sekitar.

Persajakan berhubungan dengan ritme dan intonasi. Fonologi menganalisis bunyi yang tergabungkan untuk membentuk bahasa. Morfologi menyimak komponen-komponen (morfem) yang memiliki makna yang membentuk kata, antara lain melibatkan awalan dan akhiran yang mengubah arti kata-kata akar. Sintaks merupakan studi yang menyimak tentang aturan penggabungan kata menjadi frase dan kalimat yang diperkenankan dan pemakaian aturan-aturan tersebut menjadi kalimat. Semantik menyimak arti kata, frase dan kalimat, serta cara penyampaiannya dalam bahasa alami. Pragmatik merupakan studi tentang cara penggunaan bahasa dan pengaruhnya pada pendengar.

Pengetahuan tentang dunia sekitar mencakup pengetahuan tentang dunia fisis, dunia interaksi sosial, dan peranan tujuan dalam komunikasi.

4. POLA KALIMAT BAHASA INDONESIA

Pola kalimat bahasa Indonesia yang akan dipaparkan disini dibatasi hanya pada pola kalimat yang berkaitan dengan permintaan menunjukkan data atau pertanyaan terhadap suatu informasi yang terdapat dalam suatu bahasa basis data.

Kalimat permintaan dapat diekspresikan sebagai kalimat perintah biasanya dimulai dengan kata (*verb*) yang tidak berawalan, seringkali juga predikat yang menyatakan perintah ditandai oleh partikel *-lah* [6]. Sedangkan pertanyaan terhadap suatu data dapat diekspresikan sebagai kalimat tanya yang diawali dengan kata tanya, seperti apa, siapa, berapa, bagaimana dan sebagainya [6].

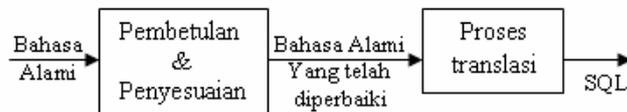
5. METODE TRANSLASI

Metode yang digunakan untuk mentranslasi bahasa alami adalah dengan melihat kata-kata yang ada pada kalimat alami tanpa melihat struktur kalimatnya. Kata-kata tersebut kemudian dibandingkan dengan daftar *keyword* untuk mendapatkan *field* yang dimaksudkan.

Selain terhadap daftar *keyword*, kata-kata alami tersebut juga dibandingkan dengan daftar kata pengantar yang berisi daftar kata-kata alami yang mungkin digunakan sebagai kriteria seperti kata-kata diawali, diakhiri, lebih dari, kurang dari, dan lainnya untuk mendapatkan kriterianya. Selanjutnya keterangan kriteria akan didapatkan dari hasil perbandingan kata-kata alami dengan daftar kata

pengantar yang berisi kata-kata pengantar keterangan kriteria seperti kata, huruf, karakter dan lain-lain.

Sebelum memasuki tahap translasi, kata-kata pada kalimat bahasa alami terlebih dahulu diperiksa untuk dibenarkan dan disesuaikan seperti Gambar 2.



Gambar 2. Diagram blok proses Pembetulan kata dan translasi

Pada tahap ini ada dua proses yang dilakukan, yaitu membenarkan kata-kata yang salah dan pengubahan bentuk karakter ke bentuk numerik.

5.1 Pembetulan dan Penyesuaian Kata

Kata-kata dalam kalimat alami tidak semuanya diperiksa untuk dibetulkan jika terjadi kesalahan, pemeriksaan hanya dilakukan terhadap kata-kata yang diperlukan untuk proses pentranslasi seperti kata-kata pada *keyword*, daftar kriteria dan daftar kata pengantar. Dengan kata lain kata-kata pada kalimat alami dibandingkan dengan kata-kata pada daftar di atas, jika terjadi kesalahan pengetikan maka kata-kata tersebut akan dibetulkan.

Beberapa kesalahan pengetikan yang mungkin terjadi seperti berikut ini:

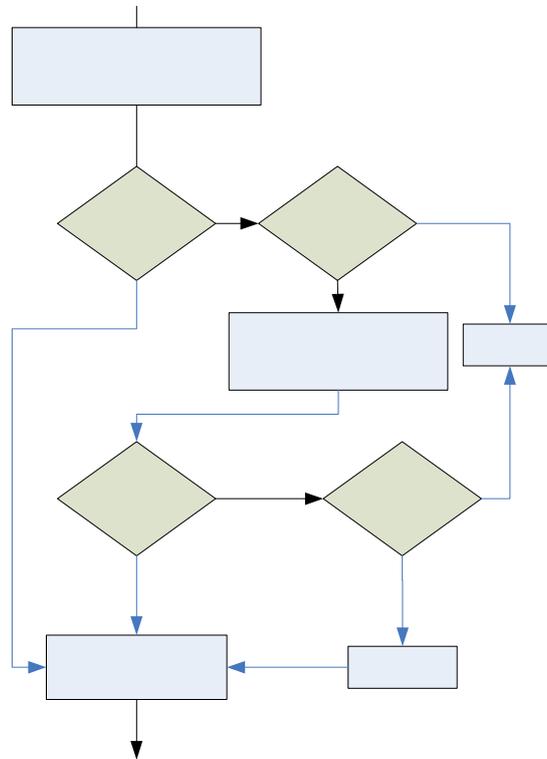
- Kekurangan huruf (contoh: Gajah ditulis Gjah)
- Kelebihan huruf (contoh: Gajah ditulis Ganjah)
- Kesalahan letak huruf (contoh Gajah ditulis Gajha)

Setiap kata dalam kalimat alami dibandingkan dengan kata-kata dalam daftar *keyword*, daftar kriteria dan daftar kata pengantar yang selanjutnya disebut sebagai kata-kata pembanding.

Karena hasil proses translasi memerlukan bentuk numerik bagi kata-kata yang menunjukkan nilai bilangan, maka kata-kata tersebut harus diubah terlebih dahulu ke bentuk numerik. Misalnya kata seribu dalam kata alami akan disesuaikan menjadi 1000.

Kata yang menunjukkan bilangan dalam bahasa Indonesia kata pertama menunjukkan bilangan puluhan dan kata terakhir menunjukkan bilangan satuan, diantara kata tersebut terdapat kata “puluh”. Misalnya dua puluh sembilan, kata pertama adalah dua (2) dan kata terakhir adalah sembilan (9), dan di antara kata-kata tersebut terdapat kata puluh. Jika satuannya sama dengan nol (0), maka kata terakhir dihilangkan.

5.2 Proses Translasi



Kalimat Alami
 Proses 1
 Penyaringan format SQL terhadap *keyword*

Gambar 3. Diagram alir proses translasi

Dari gambar diagram alir proses (Gambar 3) translasi dapat diketahui bahwa metode yang dikembangkan memiliki tiga macam metode yaitu:

1. Memilih format SQL yang sesuai dengan maksud kalimat alami
2. Menentukan kriteria format SQL yang telah dipilih.
3. Melengkapi kriteria tersebut.

Jumlah format sql SQL=1
 Tidak
 Ya
 Proses

6. PERANCANGAN SISTEM

Sistem yang dimaksud dalam tulisan ini adalah sistem pentranslasi dari bahasa alami menjadi bahasa dalam format SQL untuk proses pencarian pada suatu basis data. Sistem tersebut berupa perangkat lunak (*software*) komputer dengan input teks bahasa Indonesia sebagai bahasa alami dan output berupa SQL

Selain harus memenuhi persyaratan input dan output, sistem juga harus memenuhi syarat-syarat. Syarat-syarat sistem secara lengkap harus dapat:

1. Menerima input berupa teks bahasa Indonesia
2. Menampilkan hasil output berupa teks format SQL
3. Menampilkan tabel terhadap format SQL yang dihasilkan

Jumlah format sql SQL=1
 Format s yang mun
 229

Ya
 Proses 3
 Melengkapi format dengan kriteria

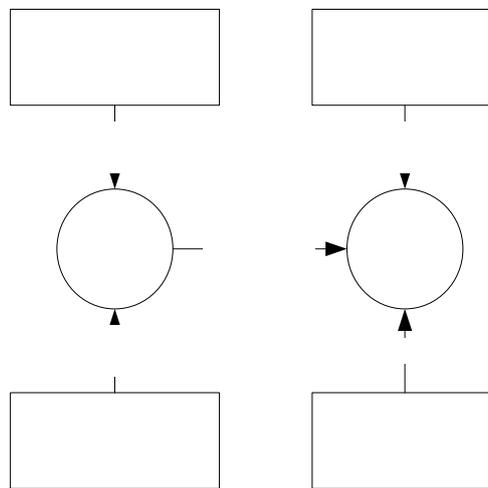
4. Menampilkan tahap-tahap proses translasi
5. mengaplikasikan basis data yang berbeda-beda.

Perancangan harus meliputi persiapan sistem yang terdiri dari persiapan tabel, penyusunan daftar kriteria, penyusunan daftar kata pengantar, dan penyusunan *keyword* terhadap format SQL serta unit-unit sistem yang terdiri dari unit antar muka, unit translasi, unit I/O file, dan unit restruktur kata.

7. IMPLEMENTASI

Sistem ini diimplementasikan pada sebuah program dengan menggunakan software Delphi yang dijalankan dengan sistem operasi Windows.

Program, ini mempunyai Data Flow Diagram (DFD) seperti pada Gambar 4.



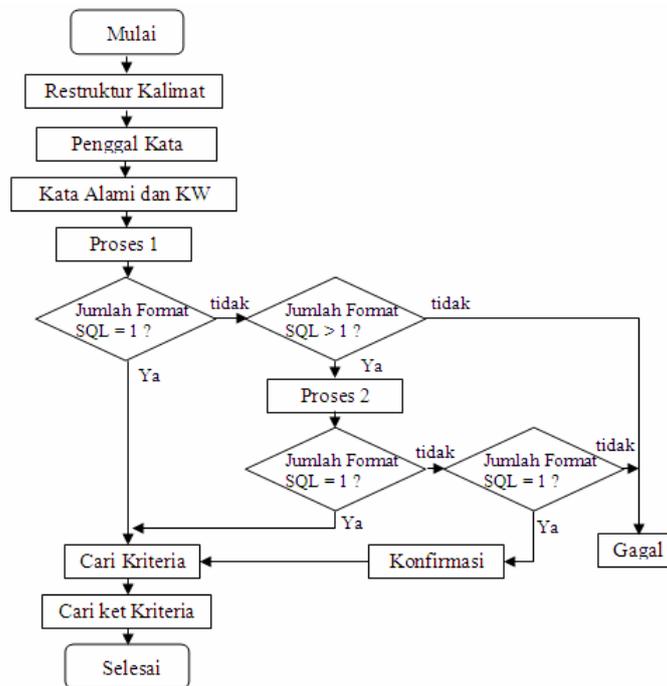
Gambar 4. Diagram alir data program

Saat mulai dijalankan program melakukan inisialisasi awal mempersiapkan sistem dan melakukan pengambilan data-data yang diperlukan pada file-file pendukung.

Pada program ini terdapat dua perintah utama yang dapat dieksekusi oleh pemakai. Kedua perintah tersebut diimplementasikan ke dalam dua prosedur yaitu "Translate" dan "Run". Translate digunakan untuk mentranslasi kalimat dalam bahasa Indonesia menjadi format SQL sedangkan prosedur Run digunakan untuk mengeksekusi format SQL dengan SQL engine terhadap tabel basis data.

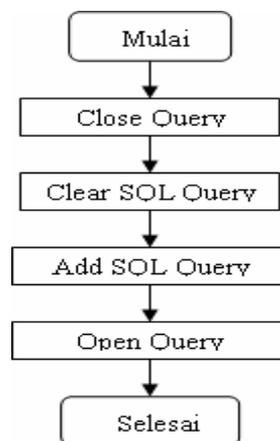
Prosedure Translate memiliki diagram alir seperti Gambar 5.

Teks Masukan



Gambar 5. Diagram alir prosedur "Translate"

Sedangkan prosedur Run memiliki diagram alir seperti pada Gambar 6.



Gambar 6. Diagram alir prosedur Run

8. PENGUJIAN METODE

Pengujian metode yang dipakai dalam perancangan sistem ini dilakukan dengan mengimplementasikan program yang menggunakan metode tersebut terhadap tiga buah basis data, yaitu Basis data akademik terdiri dari tiga tabel

(Tabel 1, 2 dan 3), basis data kendaraan terdiri dari dua tabel (Tabel 4 dan 5) dan basis data propinsi, terdiri dari satu tabel.

Tabel 1. Data akademik

<i>KODEMK</i>	<i>KODEDSN</i>	<i>PESERTA</i>
EL-111	0001	19
EL-102	0003	26
EL-201	0002	17
EL-202	0001	25
EL-301	0005	20
EL-302	0006	12

Tabel 2. Data dosen

<i>KODED</i>	<i>NAMADSN</i>	<i>GELARDSN</i>
0001	Fahmi	Dr.
0002	Faraz	Ir. M.T.
0003	Hafidz	Dr.
0004	Abdullah	Ir.
0005	Abu Bakar	Dr.
0006	Umar	Dr.

Tabel 3. Data matakuliah

<i>KODEMK</i>	<i>NAMA</i>	<i>SKS</i>	<i>SEMESTER</i>
EL-111	Kapita Selekt	2	1
EL-102	Mikroelektric	3	1
EL-104	Sistem distribusi	2	1
EL-201	Komputer paralel	3	2
EL-202	Arsitektur kmptr	2	2
EL-301	Rangkaian Logika	2	3
EL-302	Sistem Operasi	2	3

Tabel 4. Data kendaraan

<i>NAMAKEN</i>	<i>RODA</i>	<i>TENAGA</i>	<i>KECEPATAN</i>
Becak	3	1	20
Mobil	4	2	120
Sepeda motor	2	2	100
Sepeda	2	1	20
Bajaj	3	2	80
Truk	6	2	120
Delman	2	3	40

Tabel 5. Data jenis tenaga

<i>KODETEN</i>	<i>JENTENAGA</i>
1	Manusia
2	Mesin
3	Hewan

Input berupa teks kalimat dalam bahasa Indonesia, dapat dituliskan secara bebas dalam kotak teks, demikian juga output berupa teks dalam format SQL. Setelah mengetikkan kalimat dalam bahasa Indonesia, proses pentranslasiian dapat dimulai dengan mengklik tombol translate.

Pengujian metode dilakukan dengan program yang dibuat sesuai dengan metode pendeteksian kata. Pengujian ini dilakukan terhadap 50 (lima puluh) kalimat bahasa Indonesia yang berbeda untuk suatu contoh basis data akademik.

Agar kalimat uji terbagi rata diantara sepuluh jenis format SQL yang mungkin pada basis data pengujian, masing-masing jenis format diberikan lima kalimat uji. Untuk basis data kendaraan dan propinsi masing-masing dilakukan pengujian terhadap 25 dan 40 kalimat uji. Jumlah kalimat pengujian untuk setiap basis data disesuaikan berdasarkan jumlah format SQL yang mungkin untuk masing-masing basis data.

Input pengujian terdiri atas pasangan kata alami dan format SQL yang diharapkan, sedangkan outputnya adalah hasil format SQL.

$$\begin{aligned} \text{Input} &= (N_1, SI_1; N_2, SI_2; \dots; N_N, SI_N) \\ \text{Output} &= (SO_1; SO_2; \dots; SO_n) \end{aligned}$$

di mana:

N = jumlah kalimat yang diuji

SI = Kalimat alami

SI = Format SQL yang diharapkan

SO =Format SQL yang didapatkan

Pemberian nilai terhadap keakuratan metode adalah dengan melihat prosentase kebenaran hasil terhadap yang diharapkan. Kebenaran hasil berarti kalimat alami dapat ditranslasi tanpa konfirmasi dan sesuai dengan hasil yang diharapkan sedangkan kegagalan berarti tidak diperoleh satupun format SQL atau diperoleh satu format SQL tetapi tidak sesuai dengan hasil yang diharapkan, jadi penilaiannya adalah:

$$Score_n = \begin{cases} 1, SI_n = SO_n \\ 0, SI_n \neq SO_n \end{cases}$$

$$\text{Total nilainya adalah } \frac{\sum Score}{n} \times 100\%$$

Contoh hasil pengujian disajikan dalam Tabel 6.

Tabel 6. Hasil pengujian

No.	Kalimat Alami, hasil yang diharapkan dan didapatkan	P1	P2	Ket
1	Tolong carikan kode mata kuliah yang mengandung angka 1 Select kode, nama from makul where kode like "%1%" makul where kode like "%1%"	1,3,8,9,10	1	B2
2	Carikan kode mata kuliah yang mengandung angka 10 Select code, nama from makul where kode like "%10%" Select code, nama from makul where kode like "%10%"	1,3,8,9,10	1	B2
3	Mata kuliah apa saja yang namanya berakhiran huruf i select code, nama from makul where nama like "%i" select code, nama from makul where nama like "%i"	2,		B1
4	Carikan mata kuliah yang diawali huruf A select code, nama from makul where nama like "A%" select code, nama from makul where nama like "A%"	1,2,3,4,8,9,10	1,2	K
5	Tolong carikan kode dosen yang mengandung unsur angka 1 Select*from dosen where coded like "%1%" Select*from dosen where coded like "%1%"	5,8,9,10	5	B2
6	Sebutkan dosen-dosen yang bergelar Dr. Select*from dosen where gelardsn like "Dr." Select*from dosen where gelardsn like "Dr."	7		B1
7	Sebutkan kode mata kuliah yang diasuh oleh Fahmi Select kodedsn, namadsn, gelardsn, kodemk, nama, sks, from dosen, makul, akademik where kode mk = kode and koded=kodedsn and namadsn lke"Fahmi" Select kodedsn, namadsn, gelardsn, kodemk, nama, sks, from dosen, makul, akademik where kode mk = kode and koded=kodedsn and namadsn lke"Fahmi"	8		B1

Tabel 6. lanjutan.

No.	Kalimat Alami, hasil yang diharapkan dan didapatkan	P1	P2	Ket
8	Siapa nama dosen yang mengajar mata kuliah Sistem Operasi	9,10	9	B2
	Select kodemk, nama, sks, semester, namadsn, peserta from dosen, makul, akademik, where kodemk= kode and koded=kodedsn and nama like "Sistem Operasi"			
	Select kodemk, nama, sks, semester, namadsn, peserta from dosen, makul, akademik, where kodemk= kode and koded=kodedsn and nama like "Sistem Operasi"			
9	Berapa sks mata kuliah Kapita Seleкта	3, 8, 9	8,9	K
	Select kodemk, nama, sks, semester, namadsn, peserta from dosen, makul, akademik, where kodemk= kode and koded=kodedsn and nama like "Kapita Seleкта"			
	Select kodemk, nama, sks, semester, namadsn, peserta from dosen, makul, akademik, where kodemk= kode and koded=kodedsn and nama like "Kapita Seleкта"			
10	Ada pada semester berapa mata kuliah Kapita Seleкта	4, 9	4	G
	Select kodemk, nama, sks, semester, namadsn, peserta from dosen, makul, akademik, where kodemk= kode and koded=kodedsn and nama like "Kapita Seleкта"			
	Select kode, nama, semester, from makul where semester= berapa			

Keterangan:

P1 = Daftar nomor format SQL yang didapat dari proses pertama

P2 = Daftar nomor format SQL yang didapat dari proses kedua

B1 = Berhasil pada proses pertama

B2 = Berhasil pada proses kedua

K = Memerlukan konfirmasi

G = Gagal

Dari data pengujian, dengan sampel sebanyak 50 kalimat diperoleh hasil sebagai berikut:

1. 24 kalimat berhasil pada penyaringan proses satu.
2. 21 kalimat berhasil pada penyaringan proses dua
3. 4 kalimat yang memerlukan konfirmasi (memiliki lebih dari satu kemungkinan)
4. 1 kalimat gagal ditranslasi

Dari hasil yang didapatkan berarti kalimat yang berhasil ditranslasi adalah sebanyak $24+21 = 45$ kalimat atau sebesar 90% dari 50 kalimat.

Untuk pengujian basis data kendaraan, dilakukan terhadap 25 kalimat uji dengan hasil seperti pada Tabel 7.

Tabel 7. Hasil pengujian

No.	Basisdata yang diuji	P1	P2	K	G	Prosentase keberhasilan (%)
1	Akademik	24	21	4	1	90
2	Kendaraan	13	9	3	0	88
3	Propinsi	22	10	8	0	80

Dari tabel hasil pengujian tiga basis data dapat diketahui rata-rata keakuratan hasil sebesar 86%.

9. SIMPULAN

Metode pendeteksian kata yang dikembangkan pada penelitian ini dapat berfungsi untuk mentranslasikan kalimat bahasa alami menjadi format SQL untuk mencari data dengan keberhasilan rata-rata sebesar 86% terhadap tiga basis data yang digunakan dalam pengujian.

Pentranslasian dari bahasa Indonesia menjadi format SQL dapat dilakukan dengan memfokuskan pada kata-kata yang terdapat pada kalimat tanpa memperhatikan struktur kalimatnya.

PUSTAKA

- [1] Beyon, P., dan Davies. (1991) *Expert Database System: A Gentle Introduction*, McGraw-Hill.
- [2] Colburn, R. (1999) *Using SQL*, Que, Indianapolis.
- [3] Fathansyah. (1999) *Basis Data*, Informatik, Bandung.
- [4] Krulee, G.K. (1991) *Computer Processing of Natural Language*, Prentice-Hall, New Jersey.
- [5] Obermeier, K.K. (1989) *Natural Language Processing Technologies in Artificial Intelligence*, Ellis Horwood Limited, Chichester.
- [6] Sugono, D. (1999) *Berbahasa Indonesia dengan Benar*, Puspa Swara, Jakarta.
- [7] Teixeira, S., dan Xavier, P. (1999) *Delphi 5 Developer's Guide*, Sams Publishing, Indianapolis.