

**Penyusunan Aturan Produksi Bahasa Pengganti *Query*
Untuk Basis Data Matakuliah Program Studi Matematika FMIPA UNY**

Sri Andayani

Jurusan Pendidikan Matematika FMIPA UNY

Abstrak

Sebuah aplikasi pemrosesan bahasa alami dalam basis data dapat melahirkan bahasa pengganti query yang digunakan untuk mengakses informasi dari basis data tersebut. Proses pembuatan aplikasi didahului dengan penentuan tata bahasa (*grammar*) yang menjadi dasar pembentukan kalimat bahasa pengganti query, yang diwujudkan dalam sebuah aturan produksi. Penelitian ini bertujuan untuk membuat aturan produksi dari sebuah bahasa alami pengganti query yang ditujukan untuk mengakses informasi dari basis data matakuliah di program Studi Matematika FMIPA UNY.

Proses pembuatan aturan produksi diawali dengan mengidentifikasi keteraturan kalimat-kalimat pertanyaan yang biasa muncul untuk mencari sebuah informasi dari basis data matakuliah. Berdasarkan keteraturan tersebut, dibuat pengelompokan simbol-simbol bahasa yang kemudian direpresentasikan dengan menggunakan notasi *Backus Naur Form*, sehingga terbentuklah sebuah aturan produksi. Selain itu, aturan produksi yang terbentuk dapat juga disajikan dalam wujud pohon penurunan kalimat.

Aturan produksi untuk bahasa pengganti query dalam basis data matakuliah program studi Matematika diawali dengan symbol <pengganti_query> → <kata_tanya> <bagian ditanyakan> <bagian><diketahui> | <bagian ditanyakan> <kata_tanya> <bagian> <diketahui>. Symbol awal tersebut kemudian dikembangkan lebih lanjut untuk mendapatkan tata bahasa yang lebih lengkap.

Kata Kunci: **Aturan Produksi, Bahasa Pengganti *Query*, Basis Data Matakuliah**

Pendahuluan

1. Latar Belakang Masalah

Selama ini Program Studi Matematika menggunakan kurikulum yang hanya berisi daftar mata kuliah wajib dan pilihan, beserta kode mata kuliah, prasyarat serta sebaran mata kuliah tiap semester. Deskripsi silabi yang menjelaskan isi setiap mata kuliah, kompetensi yang terkandung, buku acuan yang diajarkan beserta pokok bahasannya tidak tercantum dalam buku tersebut, melainkan dalam buku yang berbeda, yakni dalam buku

silabi. Kenyataan tersebut menunjukkan tidak ada integrasi data antara kurikulum dan silabi, yang seharusnya merupakan satu kesatuan.

Hal tersebut mengakibatkan para pengguna kurikulum tidak dapat secara cepat mengetahui kompetensi setiap mata kuliah beserta informasi penting lainnya. Demikian pula benang merah keterkaitan antar mata kuliah yang menjadi dasar pemilihan mata kuliah pilihan atau penjurusan bidang kajian tetap menjadi pertanyaan mahasiswa yang tidak dapat mereka peroleh jawabannya hanya dengan melihat buku kurikulum.

Masalah tidak adanya integrasi antara data-data di buku kurikulum dan buku silabi adalah masalah yang tidak akan terjadi jika data-data tersebut dikelola dalam komputer dengan menggunakan basis data. Perkembangan teknologi komputer dalam rekayasa bidang kecerdasan buatan, khususnya bidang pemrosesan bahasa alami memungkinkan pemakai dapat mengakses informasi dari sebuah basis data dengan menggunakan bahasa mereka sehari-hari, misalnya bahasa Indonesia.

Penerapan pemrosesan bahasa alami dalam dunia pendidikan, khususnya sebagai pengganti *query* dalam basis data mata kuliah tentu akan sangat membantu praktisi pendidikan yang terkait. Pemrosesan bahasa alami (*Natural Language Processing*) merupakan bidang ilmu yang membuat komputer dapat menganalisis, memahami, atau menghasilkan bahasa alami manusia sehingga memungkinkan manusia dapat berkomunikasi dengan komputer dengan menggunakan bahasa manusia sendiri. Tujuan rekayasa dalam bidang ini antara lain untuk menterjemahkan suatu bahasa ke bahasa yang lain, menyajikan ringkasan informasi dari suatu basis data atau untuk antarmuka basis data yang 'ramah' bagi pemakai.

Bahasa memegang peranan sangat penting dalam proses komunikasi, seperti halnya komunikasi antara manusia dengan komputer. Komputer hanya memahami kode-kode dalam digit biner 0 dan 1, sedangkan manusia mempunyai bahasa komunikasi yang tidak dimengerti oleh komputer. Diperlukan sebuah program komputer yang mampu mentransfer bahasa manusia sehingga dapat dimengerti oleh komputer.

Pengaksesan data dan informasi dari sebuah basis data biasanya menggunakan bahasa dalam format khusus yang biasa disebut dengan *query*. Penggunaan bahasa alami (bahasa yang digunakan sehari-hari) sebagai pengganti *query* akan sangat membantu pemakai basis data, dikarenakan pemakai tidak perlu memikirkan format khusus yang dimiliki sebuah

query. Akan tetapi, pembuatan bahasa pengganti *query* tersebut memerlukan proses yang sangat kompleks, yang diawali dengan penentuan sintaks dan aturan produksi dari bahasa tersebut, yang kemudian melibatkan komponen-komponen pemroses bahasa seperti *scanner*, *parser*, *translator* dan *evaluator* untuk mengolahnya.

Seperti layaknya sebuah bahasa, bahasa yang digunakan untuk mengakses informasi dari basis data mempunyai sintaks tertentu. Meskipun bahasa tersebut adalah bahasa sehari-hari, akan tetapi karena hanya digunakan khusus untuk mengakses basis data, maka hanya kalimat-kalimat tertentu saja yang akan digunakan. Sintaks dari kalimat-kalimat tersebut harus dicari aturan produksinya agar dapat digunakan sebagai dasar pemrosesan komponen pemroses bahasa.

2. Rumusan masalah

Berdasarkan uraian dalam latar belakang di atas, maka masalah yang dirumuskan dalam penelitian ini adalah :

1. Bagaimanakah tabel-tabel dasar dan atribut-atributnya yang merepresentasikan basis data mata kuliah Program Studi Matematika FMIPA UNY?
2. Apa sajakah kalimat-kalimat berbahasa Indonesia yang digunakan untuk mengakses data dan informasi dari basis data tersebut?
3. Bagaimanakah aturan produksi dari sintaks bahasa alami (bahasa Indonesia) sebagai pengganti *query* untuk basis data mata kuliah Program Studi Matematika FMIPA UNY?

3. Tujuan penelitian

Penelitian yang akan dilakukan bertujuan untuk:

1. Mendefinisikan tabel-tabel dan atribut-atributnya yang merepresentasikan basis data mata kuliah Program Studi Matematika FMIPA UNY
2. Menentukan kalimat-kalimat berbahasa Indonesia yang digunakan untuk mengakses data dan informasi dari basis data tersebut.
3. Menyusun aturan produksi dari sintaks bahasa alami (bahasa Indonesia) sebagai pengganti *query* untuk basis data mata kuliah Program Studi Matematika FMIPA UNY

4. Manfaat Hasil Penelitian

Manfaat yang diharapkan dapat dipetik dari hasil penelitian ini, adalah: (1). Dapat menjadi dasar penyusunan program komputer yang berfungsi sebagai pengganti *query* untuk mengakses basis data mata kuliah Program Studi Matematika FMIPA UNY, dan (2). Menambah khasanah pengembangan ilmu komputer, khususnya dalam bidang pemrosesan bahasa alami, sehingga mendorong munculnya aplikasi sejenis yang lebih luas.

Metode Penelitian

Metode yang digunakan dalam penelitian ini meliputi langkah-langkah sebagai berikut:

1. Akan dikaji beberapa pustaka yang terkait dengan aplikasi pemrosesan bahasa alami sebagai bahasa pengganti *query* dalam basis data,
2. Mendefinisikan tabel-tabel beserta atribut-atributnya untuk merepresentasikan basis data mata kuliah Program Studi Matematika,
3. Mendefinisikan kalimat-kalimat berbahasa Indonesia yang akan sering digunakan untuk mengakses basis data tersebut,
4. Berdasarkan kalimat-kalimat tersebut, dibentuk suatu aturan produksi yang bersifat umum, sebagai dasar pemrosesan bahasa alami.

Hasil dan Pembahasan

1. Hasil Penelitian

Penyusunan aturan produksi bahasa pengganti *query* dalam basis data matakuliah di program studi Matematika diawali dengan penentuan basis data yang dijadikan obyek penelitian.

Pada dasarnya basis data adalah kumpulan tabel-tabel yang menyimpan data-data tentang matakuliah dan relevansinya. Ada 3 (tiga) buah tabel yang diperlukan agar dapat merepresentasikan kebutuhan data dan informasi tentang keterkaitan matakuliah di program studi Matematika. Tabel-tabel tersebut adalah:

- 1). Tabel Matakuliah yang mencakup atribut-atribut:
 - a. kode matakuliah, yaitu kode matakuliah
 - b. nama matakuliah, yaitu nama matakuliah

- c. sks, besarnya sks matakuliah
 - d. semester, semester diselenggarakannya matakuliah tsb.
 - e. prasyarat, matakuliah yang menjadi prasyarat
 - f. status, yakni nama kelompok matakuliah (MKK, MKU dll)
- 2). Table deskripsi yang mencakup atribut-atribut
- a. kode matakuliah
 - b. kompetensi, yakni standar kompetensi matakuliah
 - c. deskripsi, yakni deskripsi matakuliah
- 3) Tabel buku referensi yang mencakup atribut-atribut
- a. kode matakuliah
 - b. kode buku referensi
 - c. pengarang
 - d. judul
 - e. tahun
 - f. penerbit
 - g. kota

Aplikasi pemrosesan bahasa alami sebagai pengganti query basis data mendefinisikan sebuah pemakaian bahasa alami, dalam hal ini adalah bahasa Indonesia, yang digunakan untuk mengakses data dari basis data tersebut. Dalam aplikasi ini, basis data dapat diakses dengan menggunakan kalimat-kalimat sehari-hari berbahasa Indonesia. Dengan demikian, meskipun bahasa Indonesia sudah mempunyai grammar dan aturan produksi, akan tetapi dalam aplikasi ini perlu ditentukan aturan produksi yang secara khusus menangani pola-pola pertanyaan pada input.

Aturan produksi dapat ditentukan berdasarkan pola keteraturan pertanyaan untuk mengakses basis data. Keteraturan tersebut dapat diketahui berdasarkan daftar semua kalimat pertanyaan yang mungkin untuk mengakses basis data terkait.

Pertanyaan seputar informasi yang dapat diakses dari basis data dapat melibatkan satu buah tabel atau melibatkan dua atau lebih tabel. Berikut ini adalah contoh daftar pertanyaan-pertanyaan yang mungkin ditujukan dalam mengakses basis data matakuliah yang mempunyai tabel-tabel seperti tersebut di atas.

Pertanyaan yang berkisar hanya pada satu tabel

A. pada tabel matakuliah

a). Yang menanyakan kode matakuliah

1. Apakah kode matakuliah Jaringan komputer?
2. Apakah kode matakuliah Aplikasi komputer dan Analisis Regresi?
3. Apakah sajakah kode matakuliah yang depannya SMT?
4. kode matakuliah apa saja yang muncul pada semester 1?

b). Yang menanyakan nama matakuliah

1. Apakah nama matakuliah yang kodenya MAT 322?
2. Apa sajakah nama matakuliah dalam kelompok matakuliah pilihan komputer?
3. Apakah nama matakuliah yang kodenya MAT 301 dan MAT 305?
4. Apa sajakah nama matakuliah keahlian umum?
5. Apakah nama matakuliah yang prasyaratnya Kalkulus Integral?

c). Yang menanyakan sks

1. Berapa total sks matakuliah wajib?
2. Berapa jumlah sks seluruh matakuliah wajib?

d). Yang menanyakan semester

1. Semester berapa munculnya matakuliah Analisis Nyata I?
2. Di semester berapa MAT 201 muncul?

e). Yang menanyakan prasyarat

1. Apakah prasyarat untuk matakuliah Aljabar linear II?

f). Yang menanyakan status

1. Apakah Kalkulus termasuk matakuliah perilaku berkarya?

B. Pada tabel deskripsi

1. Apakah isi matakuliah Jaringan komputer?
2. Apakah kompetensi dari matakuliah Analisis Runtun waktu?
3. Apakah deskripsi matakuliah Jaringan komputer?

C. Pada tabel referensi

1. apakah judul buku karangan Hamacher?
2. Apakah penerbit buku yang kodenya Pur 3.1.12?

Pertanyaan yang melibatkan dua atau lebih tabel

1. Apakah kode matakuliah yang kompetensinya memuat kata penelitian?
2. Apakah nama matakuliah yang buku referensinya berjudul ‘Media Pembelajaran?’
3. Apakah nama matakuliah yang buku referensinya berkode HAM 4.0.2?

Pada dasarnya pertanyaan-pertanyaan tersebut terdiri dari tiga bagian, bagian pertama adalah bagian kata tanya, bagian kedua adalah bagian yang ditanyakan dan bagian ketiga adalah bagian yang diketahui. Sebagai contoh dalam pertanyaan ‘apakah deskripsi matakuliah Jaringan komputer’, bagian kata tanya adalah ‘apakah’, bagian yang ditanyakan, adalah ‘deskripsi’, sedangkan bagian yang diketahui adalah ‘Jaringan Komputer’. Secara lebih rinci, bagian-bagian tersebut mempunyai karakteristik sebagai berikut:

- i). Bagian kata tanya, yang berisi kata tanya. Misal apakah, berapa, kapan.
- ii). Bagian yang ditanyakan, selanjutnya disebut bagian ditanyakan. Bagian ini umumnya merupakan pertanyaan terhadap salah satu nama atribut dari tabel yang terdapat dalam basis data. Dalam bagian yang ditanyakan dapat terdiri dari satu atribut, misalnya kode matakuliah, nama matakuliah, semester, prasyarat, deskripsi, kompetensi, judul referensi (contoh, a – f), atau dua atribut seperti nama matakuliah dan sks, nama matakuliah dan prasyarat, nama matakuliah dan kompetensi, kompetensi dan deskripsi (contoh g – h).
- iii). Bagian yang diketahui, selanjutnya disebut bagian diketahui. Bagian ini berisi atribut dan data yang diketahui berkaitan dengan atribut pada bagian yang ditanyakan. Bagian diketahui memberi batasan kondisi dari atribut pada bagian ditanyakan. Bagian diketahui dapat terdiri dari satu kondisi atau dua kondisi diketahui. Bagian diketahui yang terdiri dari satu kondisi dalam contoh-contoh kalimat di atas adalah sebagai berikut:
 - a. untuk matakuliah Kalkulus Integral
 - b. yang kode matakuliahnya MAT 322
 - c. munculnya matakuliah Analisis Nyata
 - d. matakuliah Aljabar Linear II

Bagian diketahui yang terdiri dari dua kondisi dalam contoh-contoh kalimat di atas adalah sebagai berikut:

- a. yang merupakan matakuliah pilihan kelompok komputer
- b. apakah kode matakuliah yang kompetensinya memuat kata penelitian?

Selain itu pola tersebut di atas, ada juga beberapa kalimat yang bagian kata tanyanya muncul setelah bagian ditanyakan.

Kalimat-kalimat pertanyaan tersebut di atas membentuk sebuah keteraturan, dengan pola yang tertentu. Secara garis besar pola keteraturan yang muncul dalam kalimat-kalimat tersebut terdiri tiga bagian seperti yang telah diuraikan di atas. Berdasarkan pola keteraturan tersebut, dapat dibentuk sebuah grammar yang secara umum berlaku terhadap semua pertanyaan di atas, dan juga dapat menjadi dasar bagi pembentukan kalimat yang lain dalam konteks untuk mengakses basis data yang bersangkutan. Grammar yang dibentuk mempunyai komponen-komponen sebagai berikut:

- a. $N = \{ \langle \text{pengganti_query} \rangle, \langle \text{kata_tanya} \rangle, \langle \text{bagian_ditanyakan} \rangle, \langle \text{bagian_diketahui} \rangle, \langle \text{atribut} \rangle, \langle \text{lanjutan_atribut} \rangle, \langle \text{kata_sambung} \rangle, \langle \text{kosong} \rangle, \langle \text{atr_op_data} \rangle, \langle \text{operator} \rangle, \langle \text{operator} \rangle, \langle \text{kosong} \rangle, \langle \text{data} \rangle \}$
- b. $\Sigma = \{ \text{apakah, berapakah, berapa, apa saja, kapan, dimana, bagaimana, dan, atau, kode_mk, nama_mk, sks, semester, prasyarat, status, deskripsi, kompetensi, kode_buku, judul, pengarang, penerbit, tahun, kota, =, >, <, >=, <=, <>, } \}$
- c. P= aturan produksi, sebagai berikut:

Sebuah aturan produksi diawali dengan simbol awal <nonterminal>. Simbol awal dalam aturan produksi untuk kalimat-kalimat di atas digunakan <pengganti_query>. Aturan produksi selengkapnya adalah seperti berikut.

- | | |
|---------------------|---|
| <pengganti_query> | → <kata_tanya>< bagian_ditanyakan><bagian_diketahui > ..(1) |
| | < bagian_ditanyakan><kata_tanya><bagian_diketahui > .. (2) |
| <bagian_ditanyakan> | → <atribut> <lanjutan_atribut> ** .. (3) |
| <lanjutan_atribut> | → <kata_sambung> <bagian_ditanyakan > .. (4) |
| | <kosong> .. (5) |
| <bagian_diketahui> | → <atribut> <atr_op_data> <bagian_diketahui>** .. (6) |
| | <kata_sambung> <atr_op_data> < bagian_diketahui >** ..(7) |
| | <atr_op_data> < bagian_diketahui >** ... (8) |
| | bukan <atr_op_data> < bagian_diketahui >** ... (9) |
| | <kosong> ... (10) |
| <atr_op_data> | → <atribut> <operator> <data> ... (11) |
| <kata_tanya> | →apakah berapakah berapa apa saja kapan dimana bagaimana ... (12) |

<kata sambung>	→ dan atau	...(13)
<atribut>	→ kode_mk nama_mk sks semester prasyarat status deskripsi kompetensi kode_buku judul pengarang penerbit tahun kota	...(14)
<operator>	→ = > < >= <= <	...(15)
<kosong>	→	...(16)
<data>	→ tergantung dari data yang ada pada query.	...(17)

Arti notasi yang digunakan dalam aturan produksi tersebut :

→	didefinisikan sebagai
<>	simbol nonterminal
	atau
**	simbol di depannya dapat diulang nol atau satu kali
/ # # /	keterangan

2. Pembahasan

Aturan produksi menggunakan notasi-notasi yang berdasarkan pada aturan *Backus Naur Form*. Pada dasarnya aturan penulisan symbol-simbol dalam aturan produksi adalah sebagai berikut :

→	didefinisikan sebagai
<>	simbol nonterminal
	atau
**	simbol di depannya dapat diulang nol atau satu kali
/ # # /	keterangan

Aturan produksi yang telah diperoleh dapat digunakan untuk menurunkan kalimat yang secara umum digunakan dalam pengaksesan informasi dari basis data matakuliah.

Penurunan kalimat dengan menggunakan aturan produksi tersebut di atas diilustrasikan dengan menggunakan contoh berikut:

Misal pertanyaan :

Apakah nama matakuliah yang buku referensinya berkode HAM 4.0.2?

dapat diperoleh dari aturan produksi tersebut dengan langkah-langkah:

<pengganti_query> → <kata_tanya><bagian_ditanyakan><bagian_diketahui>

dengan mengganti <bagian_ditanyakan> dengan aturan produksi nomor (3) dan mengganti <bagian_diketahui> dengan aturan produksi nomor (6) diperoleh

<kata_tanya><atribut><lanjutan_atribut> ** <atribut> <atr_op_data> <bagian_diketahui>**

Selanjutnya, dengan menerapkan pengertian **, bahwa simbol di depannya dapat diulang nol atau satu kali, maka <lanjutan_ atribut> dan <bagian_diketahui> bisa dianggap diulang nol kali. Berdasarkan hal tersebut diperoleh :

<kata_tanya> <atribut> <lanjutan_ atribut> <atribut> <atr_op_data> <bagian_diketahui>

Kemudian dengan menggunakan aturan produksi nomor (5) untuk <lanjutan_ atribut> dan aturan produksi nomor (10) untuk <bagian_diketahui>, menyebabkan simbol-simbol tersebut menjadi <kosong> (tidak ada). Dengan demikian, diperoleh:

<kata_tanya> <atribut> <atribut> <atr_op_data>

Terakhir, dengan menerapkan aturan produksi nomor (11), dengan memilih simbol <data> diperoleh:

<kata_tanya> <atribut> <atribut> <data>

Dengan mengganti masing-masing simbol non terminal dengan simbol terminal yang diinginkan, diperoleh

<kata_tanya> → apakah

<atribut> → nama matakuliah

<atribut> → kode_buku

<data> → HAM 4.0.2?

Manfaat aturan produksi

Aturan produksi menjadi dasar bagi pembuatan aplikasi pemrosesan bahasa alami sebagai bahasa pengganti query untuk mengakses informasi dalam basis data matakuliah. Pada tahap selanjutnya, setelah dibuat aturan produksi perlu dibuat sebuah *parser*, yang merupakan bentuk dari *analytic grammar*. *Parser* akan mengevaluasi apakah sebuah kalimat yang diinput merupakan kalimat yang bisa diterima oleh aplikasi pemrosesan bahasa alami yang dikembangkan berdasarkan aturan produksi di atas.

Simpulan Dan Saran

1. Simpulan

Basis data Matakuliah yang digunakan dalam penyusunan aturan produksi terdiri dari 3 buah tabel, yaitu Matakuliah, Deskripsi dan Buku Referensi. Tabel Matakuliah terdiri dari dengan atribut-atribut kode matakuliah, nama matakuliah, sks, semester, prasyarat, dan status. Tabel Deskripsi terdiri dari atribut-atribut kode matakuliah, kompetensi dan deskripsi. Sedangkan tabel Buku Referensi terdiri dari atribut-atribut kode matakuliah, kode buku referensi, pengarang, judul, tahun, penerbit dan kota.

Kalimat-kalimat yang digunakan untuk mengakses data dari basis data tersebut pada dasarnya terdiri dari tiga bagian yaitu bagian kata tanya, bagian yang ditanyakan, dan bagian yang diketahui. Berdasarkan ketereaturan pola dalam kalimat-kalimat tersebut, dapat dibentuk sebuah *grammar*. Grammar yang dibentuk terdiri dari himpunan simbol nonterminal N , simbol terminal Σ dan serangkaian aturan produksi P . Aturan produksi diawali dengan symbol awal <pengganti_query> .

2. Saran

Aplikasi pemrosesan bahasa alami (*Natural Language Processing*) sebagai pengganti *query* dalam basis data Matakuliah di Program Studi Matematika FMIPA UNY membutuhkan proses yang sangat panjang dan kompleks. Aplikasi tersebut membutuhkan komponen-komponen pemroses bahasa antara lain *parser*, *scanner*, *evaluator*. Hasil penelitian ini sangat menentukan dalam pembuatan komponen pemroses bahasa alami, yaitu parser, maka disarankan ada penelitian lanjutan yang mengimplementasikan komponen pemroses bahasa alami tersebut.

DAFTAR PUSTAKA

- Carlo, G. and Jazayeri, M. 1987. *Programming Language Concepts*. Singapore : John Wiley & Sons, Inc.
- Firrar Utdirartatmo. 2001. *Teknik Kompilasi*. Yogyakarta : J & J Learning.
- Kaplan, R.M. 1994. *Constructing Language Processors for Little Languages*. New York: John Wiley & Sons, Inc.
- Kazutomo, N. , Naohito, I., Shozo, K., Chikako, K., Katsushi, S., Ikuko, T., Yoshiko, I., 1993. *Database Retrieval System for responding to Natural Language Queries with Corresponding Tables*. <http://swpat.ffii.org/patents/samples/ep52259/index.en.html>
- Maier, D. and Warren, D.S. 1988. *Computing with Logic. Logic programming with Prolog*. Menlo Park : The Benjamin/Cummings Publishing Company Inc.
- Martin, J.C. 1997. *Introduction to languages and theory of computation*. Singapore : McGraw–Hill Book Co.
- Ralston, A. and Reilly, E.D. 1993. *Encyclopedia of Computer Science*. New York : Van Nostrand Reinhold.
- Suparman. 1991. *Mengenal Artificial Intelligence*. Yogyakarta : Penerbit Andi Offset.

_, diakses Maret 2006. *Formal grammar*. (http://en.wikipedia.org/wiki/Production_rule).