

PENERAPAN ANTAR MUKA BAHASA ALAMI DALAM Pencarian INFORMASI SKRIPSI PADA SUATU PROGRAM STUDI

Dewi Soyusiawaty¹⁾, Anna Hendri Soleliza Jones²⁾

^{1, 2)} Fakultas Teknologi Industri Program Studi Teknik Informatika Universitas Ahmad Dahlan Yogyakarta
Kampus 3 UAD Jl. Prof. Dr. Soepomo Janturan, Yogyakarta
Email : dewi.soyusiawaty@tif.uad.ac.id¹⁾, annahendri@tif.uad.ac.id²⁾

Abstrak

Data daftar judul skripsi yang saat ini didokumentasi melalui digilib.uad.ac.id belum dapat dimanfaatkan secara optimal untuk membantu mahasiswa melihat daftar judul skripsi yang sudah ada sebagai referensi dalam menentukan topik skripsi yang akan dilakukan. Data yang ditampilkan hanya berupa daftar judul sesuai kata kunci yang diinputkan. Belum dapat ditampilkan daftar judul berdasarkan topik skripsi, metode yang digunakan, objek penelitian dan tahun selesai skripsi. Penelitian ini akan menggunakan data skripsi untuk dikelola sehingga dapat memberikan informasi yang bermanfaat untuk memberikan referensi judul skripsi yang pernah ada sebelumnya. Pengambilan data dari sebuah sistem basis data relasional dapat dilakukan dengan menggunakan pemrosesan bahasa alami (Natural Language Processing/NLP). Penelitian ini bertujuan untuk membangun sistem pencarian informasi skripsi dengan antarmuka bahasa alami. Masukan berupa teks atau kalimat dalam bahasa alami akan digunakan untuk melakukan query data skripsi yang tersimpan. Pembangunan aturan produksi yang diperlukan untuk menerjemahkan bahasa alami ke dalam query terhadap basis data merupakan hal yang penting. Kebenaran kalimat yang diberikan akan diperiksa dengan melakukan analisis sintaks (parser). Kalimat yang mampu melewati tahapan parser akan diterjemahkan ke dalam bahasa query yakni SQL. Evaluasi hasil dari bahasa SQL oleh evaluator diperlukan untuk melakukan pengecekan apakah hasil sudah sesuai dengan kaidah yang dibenarkan. Hasil keluaran sistem adalah informasi skripsi yang dapat dipergunakan oleh mahasiswa atau pihak lain yang berkepentingan seperti pengelola skripsi dan pihak program studi yang ingin mencari referensi tentang judul skripsi yang sudah ada. Informasi skripsi berupa daftar judul beserta jumlah berdasar topik yang diinputkan, metode penyelesaian, objek penelitian dan tahun penyelesaian. Pertanyaan yang diajukan dalam aplikasi ini masih terbatas dan disesuaikan dengan aturan produksi yang dibangun. Model pencarian informasi dengan antarmuka bahasa alami ini menjadi sebuah alternatif pada proses pencarian informasi skripsi dalam rangka menyediakan sebuah sistem yang lebih luwes dalam berinteraksi dengan penggunaanya.

Kata kunci: bahasa alami, query, skripsi

1. Pendahuluan

Penentuan topik skripsi merupakan masalah awal yang dialami mahasiswa tingkat akhir. Mahasiswa biasanya kesulitan, ragu bahkan tidak tahu untuk menetapkan topik sehingga memerlukan waktu lebih untuk memulai skripsi. Salah satu fasilitas yang bisa dimanfaatkan mahasiswa untuk mencari skripsi yang sudah ada yaitu melalui web perpustakaan kampus. Informasi yang ditampilkan berupa judul skripsi, penulis, program studi, dan lokasi skripsi tersebut disimpan. Sistem belum dapat memfilter data berdasarkan program studi untuk suatu skripsi dengan kata kunci yang diinput. Sistem juga belum dapat memfilter judul skripsi berdasarkan topik dan tahun tertentu.

Pengambilan kembali informasi yang tersimpan dalam basis data bisa dilaksanakan dengan menggunakan pengolahan bahasa alami. Beberapa teknik dan pendekatan digunakan serta dikembangkan dengan tujuan agar komputer lebih mampu memahami instruksi/permintaan manusia melalui penggunaan bahasa sehari-hari. Bahasa alami banyak digunakan oleh sebagian orang karena kemudahan dan keluwesan dalam menentukan kata kunci penelusuran dari informasi yang dicari. Dengan kemudahan dan spesifikasi yang tinggi dalam merumuskan sebuah kata kunci maka bahasa alami dapat melakukan penelusuran informasi lebih efektif. [1]

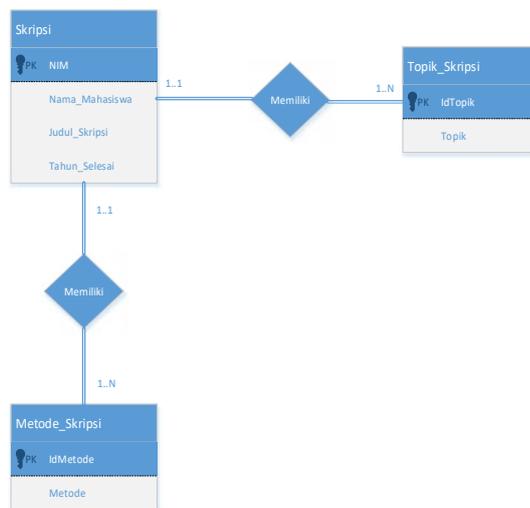
Untuk itu diperlukan aturan produksi yang secara khusus akan mengenali model pola kalimat/pertanyaan yang diberikan dari masukan sistem. Aturan produksi ditetapkan berdasarkan pola keteraturan kalimat/pertanyaan yang digunakan untuk mengakses basis data. Keteraturan pola pertanyaan dikenali dengan mendaftarkan seluruh kalimat-kalimat pertanyaan yang mungkin dapat diberikan dalam pengaksesan basis data. [2]

Penelitian menggunakan data skripsi untuk dikelola agar dapat dimanfaatkan banyak pihak baik bagi mahasiswa atau pengelola skripsi atau pengelola program studi untuk menganalisis data tersimpan sehingga dapat memberikan informasi yang bermanfaat untuk menyelesaikan permasalahan kesulitan menentukan topik dan juga lama penyelesaian skripsi. Pengambilan data dari sebuah sistem basis data relasional dapat dilakukan dengan menggunakan Natural Language Processing (NLP) dengan menggunakan bahasa alami untuk keluwesan interaksi. [3]

2. Pembahasan

2.1 Konsep Bahasa Alami untuk Pencarian Informasi Skripsi

Penelitian ini mengembangkan sistem pengolah bahasa alami dengan bahasa Indonesia untuk mendapatkan data dan informasi dari sistem basis data. Informasi yang didapatkan adalah data-data yang berkaitan dengan skripsi pada suatu program studi khususnya dalam hal ini di program studi Teknik Informatika. Pada gambar 1 di bawah merupakan perancangan basis data untuk pencarian data skripsi. Data skripsi meliputi identitas mahasiswa yang berupa NIM dan Nama dan selanjutnya judul skripsi serta tahun selesai. Untuk memberikan pengelompokan judul skripsi, maka dibentuk entitas topik skripsi dan metode skripsi.



Gambar 1. Perancangan Basis Data

Penelitian ini secara garis umumnya melakukan perancangan dan implementasi aplikasi yang menterjemahkan bahasa alami dengan bahasa Indonesia ke dalam bahasa SQL yang kemudian ditampilkan hasil dari querynya.

2.2 Aturan Produksi

Bahasa Indonesia telah memiliki grammar dan aturan produksi namun dalam penggunaannya untuk melakukan query basis data perlu diberikan aturan produksi secara khusus. Pola pertanyaan yang diberikan dalam bahasa alami untuk mendapatkan hasil dari query sistem basis data perlu disusun ke dalam suatu aturan produksi. Aturan produksi dibangun dengan melakukan pemetaan keteraturan pola dari suatu pertanyaan. [4]

Pada penelitian ini dilakukan pembatasan pertanyaan yang diijinkan untuk diinputkan yakni :

- a. Tampilkan/Daftar/Tunjukkan/dan lain-lain
Input kalimat ini untuk mencari tahu tentang daftar judul skripsi terkait dengan topik dan metode tertentu yang diinputkan. Adapun contoh kalimat inputannya tidak dibatasi dengan kata apapun, kecuali terkait topik

dan metode skripsi yang diinginkan. Berikut beberapa contoh kalimat inputan :

- 1) Tampilkan daftar judul skripsi dengan topik sistem pakar
 - 2) Daftar judul skripsi terkait pemrosesan bahasa alami
 - 3) Tunjukkan judul-judul skripsi sistem informasi geografis
 - 4) Contoh skripsi tentang multimedia
 - 5) Tampilkan contoh skripsi pemrosesan bahasa alami dengan top down parang
 - 6) Pemrosesan bahasa alami, top down parsing
 - 7) Dan lain-lain
- b. Berapa jumlah skripsi dengan topik sistem pakar?
Input kalimat berupa kata 'Berapa' pada awal kalimat digunakan untuk mengetahui jumlah dari skripsi yang ada terkait topik tertentu. Informasi jumlah skripsi diperlukan bagi koordinator skripsi dan atau pejabat program studi untuk mengetahui statistik atau perbandingan jumlah antar satu topik skripsi dengan topik lainnya. Contoh kalimat inputan :
- (1) Berapa jumlah judul skripsi tentang multimedia
 - (2) Berapa jumlah skripsi dengan topik sistem pendukung keputusan

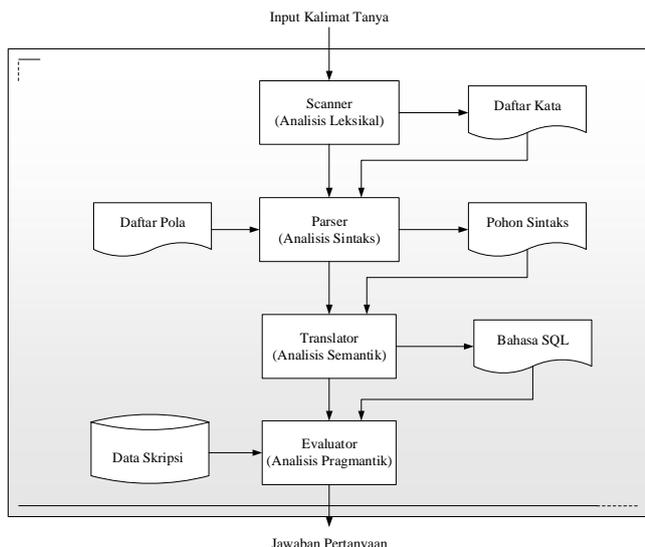
Pertanyaan – pertanyaan diatas memiliki pola keteraturan dan dapat disusun dengan notasi aturan sebagai berikut :

- a. S <topik> | <atribut> <topik> | <atribut> <topik> <atribut>
- b. S → <metode> | <atribut> <metode> | <atribut> <metode> <atribut>
- c. S → <topik> <metode> | <atribut> <topik> <metode> | <atribut> <topik> <metode> <atribut>
- d. S → <berapa> <topik> | <berapa> <metode>
- e. <atribut> → {semua kata selain topik dan metode}
- f. <topik> → {Pemrosesan Bahasa Alami, Sistem Pakar, Sistem Pendukung Keputusan, Multimedia, Sistem Informasi Geografis, dan lain-lain}
- g. <metode> → {top down parsing, restful, fuzzy tsukamoto, dan lain-lain}

Sistem yang dirancang hanya dapat menerima pertanyaan sesuai dengan aturan produksi yang ditetapkan di atas. Pertanyaan yang tidak sesuai dengan pola aturan diatas akan pesan kesalahan atau diabaikan.

2.3 Perancangan Flowchart Pemrosesan

Komponen bahasa alami pada pencarian skripsi tertera pada gambar 2 di bawah. Sebelum mendapat hasil maka setiap input teks akan melewati beberapa proses yaitu analisis leksikal, sintak, semantik, dan pragmatik. [5][6]



Gambar 2. Tahapan Pemrosesan Bahasa Alami pada Pencarian Skripsi

(1) Analisis Leksikal (Scanner)

Pemberian batasan pertanyaan dilakukan oleh subsistem yang bernama *scanner* dan *parser*. Pemrosesan awal pertanyaan dilakukan oleh *scanner* yang akan membuat pertanyaan menjadi sebuah daftar token. Pada penelitian ini *scanner* melakukan 3 aktivitas yakni: menghilangkan tanda baca, memecah pertanyaan dan mengubah menjadi huruf kecil. Pada Gambar 3 di bawah menampilkan tahap analisis leksikal dengan contoh satu input.



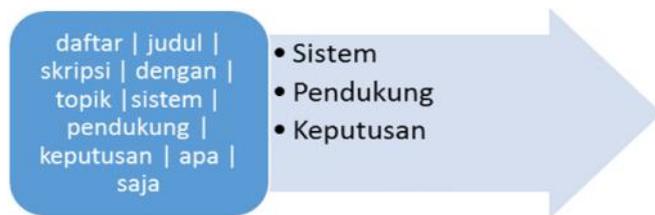
Gambar 3. Tahap Analisis Leksikal

(2) Analisis Sintaks (Parser)

Daftar token yang terbentuk akan dianalisa oleh *parser* apakah sesuai dengan dengan pembentukan pola kalimat yang telah ditetapkan. Penentuan struktur kalimat bisa jadi merupakan pekerjaan yang sulit bergantung dari bahasa yang digunakan. Parsing morfologi merupakan kegiatan yang dilakukan untuk menentukan struktur kata. Pada penelitian ini secara umum *parser* akan melakukan:

- a. Membaca daftar token
- b. Membaca cacah token
- c. Membaca isi token
- d. Membandingkan dengan aturan produksi

Analisis sintaks melakukan pelacakan terhadap token-token yang dihasilkan oleh *scanner* kemudian dibandingkan dengan daftar token yang tersedia. Jika sesuai dengan daftar token yang ada maka dilihat apakah terdapat kecocokan dengan aturan produksi yang ada. Sebagai contoh gambar 4 merupakan tahap analisis sintak dengan membaca isi token dalam termin topik apakah sesuai dengan daftar token yang diijinkan.



Gambar 4. Tahap Analisis Sintak

(3) Analisis Semantik (Translator)

Translator berfungsi untuk memetakan hasil *parser* yang sesuai dengan aturan produksi ke dalam bahasa hasil yakni bahasa *query*, sehingga yang dilakukan pada semantik meliputi :

- a. Membaca hasil scanner
- b. Memetakan ke dalam Bahasa Query

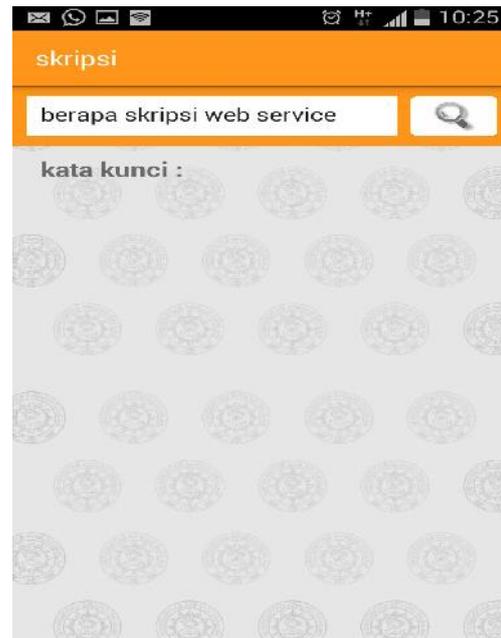
Pada tahap ini atribut dan obyek memegang peranan penting dalam translator. Sebagai contoh kalimat input “Daftar judul skripsi dengan topik sistem pendukung keputusan apa saja?”, maka dalam translator menjadi “sistem pendukung keputusan“. Kalimat tersebut akan diterjemahkan ke dalam bahasa query menjadi “Select judul from skripsi where topik = ‘sistem pendukung keputusan’.

2.4 Implementasi Sistem

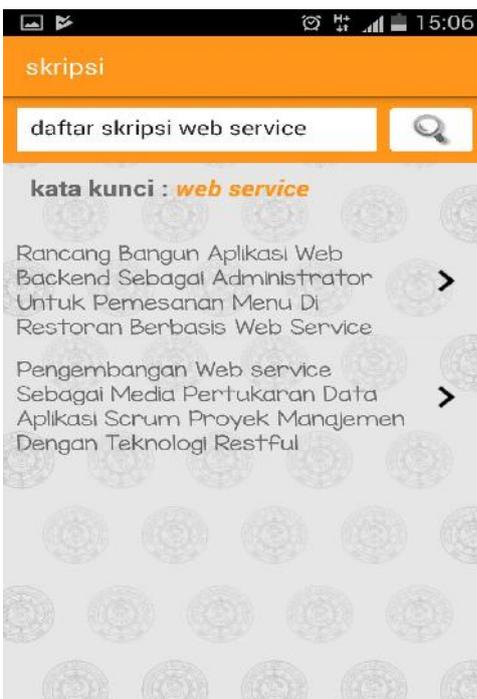
Berikut ditampilkan implementasi sistem dari tahapan pemrosesan sebelumnya. User menginputkan kalimat pencarian untuk data skripsi yang diperlukan. Hal ini ditunjukkan pada gambar 5. Dalam penelitian ini user dapat menginputkan kalimat dengan bahasa sehari-hari yang digunakan tanpa ada batasan. Selanjutnya setelah kalimat diinputkan maka akan dilakukan analisis leksikal seperti yang diuraikan di atas. Hasil dari analisis leksikal yaitu berupa daftar kata atau token. Selanjutnya akan dilanjutkan dengan tahapan analisis sintak pada gambar 6. Hasil dari analisis sintak berupa kata kunci yang diambil hasil parsing dan melihat ketersediaan kata tersebut di basis data. Pada contoh yang ada di Gambar 5, kalimat yang diinputkan adalah “daftar skripsi web service”, maka pada tahapan analisis leksikal, sistem akan mendapatkan 4 token/kata. Selanjutnya pada tahap analisis sintak didapat kata kunci ‘web service’. Pada tahap analisis semantik akan dipetakan hasil analisis sintak yang berupa kata kunci ke dalam query.



Gambar 5. Implementasi Halaman Input



Gambar 7. Pencarian berdasarkan Jumlah



Gambar 6. Pencarian berdasar kata kunci



Gambar 8. Hasil Pencarian Jumlah

Pada Gambar 5, user menginputkan data kalimat pencarian yang dapat ditulis dalam bahasa sehari-hari untuk kemudahan user dalam mencari informasi ke dalam sistem. Pada Gambar 6, merupakan hasil olahan dari sistem setelah mendapatkan input user.

Selain menampilkan judul skripsi yang tersimpan dalam basis data, sistem juga dapat mengolah kalimat pencarian untuk menampilkan jumlah data yang tersimpan. Contoh pada Gambar 7 diinputkan kalimat 'berapa skripsi web service' dengan maksud untuk menghitung berapa jumlah judul skripsi dengan topik web service. Gambar 8 menampilkan hasil dari kalimat input setelah melewati

beberapa tahap sampai dengan query yang menghasilkan jumlah baris tersimpan dalam basis data.

2.5 Pengujian

Pada pengujian dilakukan terhadap pengguna khususnya para mahasiswa, didapatkan hasil lebih dari 90% menyatakan setuju dan manfaat terhadap sistem yang dibuat. Pengujian sistem secara teknis juga sudah dilakukan kepada koordinator skripsi untuk memvalidasi fungsi-fungsi yang didefinisikan di sistem agar benar-benar dapat berfungsi dengan baik termasuk pula kebenaran dan kelengkapan isi dari basis data skripsi tersebut.

3. Kesimpulan

Penelitian ini menghasilkan sebuah aplikasi yang mampu melakukan *query* terhadap data yang disimpan dalam basis data relasional dengan menggunakan bahasa Indonesia. Kalimat pencarian yang diajukan oleh user hanya dapat diproses sesuai dengan aturan produksi yang dibangun. Pertanyaan yang tidak sesuai dengan aturan produksi yang ditetapkan akan diberikan pesan kesalahan atau diabaikan. Membebaskan pertanyaan menjadi lebih luas adalah saran yang bisa dilakukan untuk penelitian selanjutnya sehingga pengguna dapat memiliki variasi pertanyaan yang lebih banyak.

Daftar Pustaka

- [1] Lewis, D.D dan Jones, K.S. 1996, "Natural language processing for information retrieval," *Communications of the ACM*, vol. 39, no. 1, pp. 92–101, 1996.
- [2] Raharjo, Suwanto., Hartati, Sri, "Antar muka Bahasa Alami untuk Melakukan Query terhadap Terjemahan Al-Quran", *Jurnal Teknologi*, Volume 7 Nomor 1, Juni 2014.
- [3] Popescu, A.-M. Etzioni, O. dan Kautz, H, "Towards a theory of natural language interfaces to databases," in *Proceedings of the 8th international conference on Intelligent user interfaces*. ACM, pp. 149–157, 2003.
- [4] Hartati, S dan Zuliarso, E, Aplikasi pengolahan bahasa alami untuk query basis data XML, e-paper dari *Dinamik Jurnal Teknologi Informasi*, Vol. 13, No. 2, 2008.
- [5] Utami, E. dan Hartati, S. 2008, "Pendekatan Metode Rule Based dalam Mengalihbahasakan Teks Bahasa Inggris Ke Teks Bahasa Indonesia," *Jurnal Informatika*, vol. 8, no. 1, pp. pp–42, 2007.
- [6] Kao, A., & Poteet, S.R, *Natural Language Processing and Text Mining*, London: Springer-Verlag, 2007.

Biodata Penulis

Dewi Soyusiawaty, memperoleh gelar Sarjana Teknik (S.T.), Jurusan Teknik Informatika STT Telkom Bandung, lulus tahun 1999. Memperoleh gelar Magister Teknik (M.T.) Program Pasca Sarjana Universitas Gajah Mada Yogyakarta, lulus tahun 2003. Saat ini menjadi Dosen di Program Studi Teknik Informatika Universitas Ahmad Dahlan Yogyakarta.

Anna Hendri Soleliza Jones, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Ilmu Komputer Universitas Gajah Mada Yogyakarta, lulus tahun 2008. Memperoleh gelar Magister Komputer (M.Cs) Program Pasca Sarjana Ilmu Komputer Universitas Gajah Mada Yogyakarta, lulus tahun 2016. Saat ini menjadi Dosen di Program Studi Teknik Informatika Universitas Ahmad Dahlan Yogyakarta.

