

Parsing dan Konversi Kalimat Tanya Konfirmatif Menjadi Query Sparql Menggunakan Pendekatan Top-Down Parsing

Mohammad Syarief

Prodi Manajemen Informatika, Universitas Trunojoyo Madura
mohammad.syarief@trunojoyo.ac.id

ABSTRAK

Penelitian ini merupakan lanjutan penelitian sebelumnya (Syarief, 2014), yaitu dengan menambahkan fitur pengenalan kalimat tanya konfirmasi, yaitu kalimat tanya yang hanya membutuhkan jawaban ya atau tidak. Data yang dibahas dalam penelitian ini menggunakan ontologi OntoMotif (Syarief, 2015) sebagai ontologi utama. Pendekatan teknik parsing yang digunakan adalah top-down parsing dengan menekankan pada pengenalan predikat. Penyusunan gramatika kalimat tanya konfirmatif dalam penelitian ini difokuskan pada pengenalan kalimat berita sederhana berbentuk SPO (subjek – predikat - objek) dengan 4 macam predikat utama, yaitu diproduksi, memproduksi, mempunyai, dan dimiliki. Dari hasil pengujian, dapat disimpulkan bahwa pendekatan teknik parsing dan konversi dalam penelitian ini berhasil menjawab pertanyaan yang cukup bervariasi seputar fitur-fitur kendaraan bermotor.

Kata Kunci: parsing, pengenalan kalimat, kalimat tanya konfirmatif, query sparql, top-down parsing, pengenalan predikat.

PARSING AND CONVERSION OF YES-NO QUESTION INTO SPARQL QUERY USING TOP-DOWN PARSING APPROACH

ABSTRACT

This paper is a continuation of our previous paper (Syarief, 2014), by adding a feature of yes-no questions (interrogative sentences that only requires a yes or no answer) recognition. The data use OntoMotif (Syarief, 2015) as main ontology. Parsing technique approach used in this paper is top-down parsing. This approach is developed based on predicate-argument recognition of a simple sentence. The grammar focused on recognition of a simple SPO (subject - predicate - object) sentences with 4 main predicates, i.e diproduksi (manufactured), memproduksi (produced), mempunyai (have), and dimiliki (owned). Based on test results, it can be concluded that the this approach managed to answer varied questions around otomotives and their features.

Keywords: parsing, sentence recognition, yes-no question, sparql query, top-down parsing, predicate-argument recognition.

PENDAHULUAN

Transportasi merupakan bagian hidup yang tak terpisahkan dari kehidupan manusia, maka informasi mengenai kendaraan bermotor sebagai sarana transportasi tentu juga sangat diperlukan. Pada era digital ini, pencarian informasi biasanya dilakukan dengan internet sebagai media utamanya.

Salah satu alternatif pencarian informasi yang lebih terarah dari pada mesin pencari adalah sistem tanya jawab. Mesin pencari akan memberikan alternatif-alternatif jawaban dengan menyertakan link-link terkait, sedangkan sistem tanya jawab akan langsung memberikan jawaban dari pertanyaan yang diajukan (Cody Kwok, 2001).

Implementasi sistem tanya jawab mengharuskan adanya pemahaman terhadap kalimat input dalam rangka memberikan jawaban yang tepat sesuai yang diinginkan oleh penanya. Penelitian-penelitian tentang pengenalan kalimat berbahasa Indonesia sudah banyak dilakukan, diantaranya pengenalan kalimat tanya yang mengandung unsur perbandingan (Bendi, 2009). Selain itu, pemrosesan kalimat kompleks dan kreatif juga telah dikembangkan oleh (Larasati, 2007; Mahendra, 2008; Suryawan, 2013).

Penggunaan kata umum dalam pencarian sering kali menjadi sebab pencarian kurang terarah. Contohnya pencarian dengan kata kunci (kalimat) "Adakah kendaraan yang berkarburasi injeksi" akan menyebabkan hasil pencarian yang kurang valid jika sistem tidak dapat mendeteksi bahwa Mio, Vario, dan semisalnya termasuk kategori motor, dan motor merupakan salah satu

jenis kendaraan. Salah satu cara mengatasi hal tersebut adalah dengan menggunakan teknologi Web Semantik (Berners-Lee, 2001).

Penggunaan teknologi web semantik menggunakan sistem tanya jawab pada domain otomotif (SiTaSO) sudah dikembangkan sebelumnya (Syarief, 2014). Namun, SiTaSO hanya dapat menjawab pertanyaan yang menggunakan kata tanya apa, berapa, dan dimana. Dalam pendekatan pengenalan kalimatnya, SiTaSO menggunakan pendekatan pengenalan predikat. Predikat-predikat yang dapat dikenali dibatasi dengan diproduksi, memproduksi, mempunyai, dimiliki, dijual, menjual, menyewakan, dan disewakan.

Penelitian ini merupakan lanjutan penelitian sebelumnya (Syarief, 2014), yaitu dengan menambahkan fitur pengenalan kalimat tanya konfirmasi, yaitu kalimat tanya yang hanya membutuhkan jawaban ya atau tidak (*yes-no question*). Data yang dibahas dalam penelitian ini menggunakan ontologi OntoMotif (Syarief, 2015)

METODE PENELITIAN

Penelitian ini membahas tentang bagaimana cara mengenali kalimat tanya konfirmatif dan mentransformasikan pemahaman kalimat itu menjadi bentuk query SPARQL. Kombinasi kalimat tanya konfirmatif berbahasa Indonesia sangatlah banyak. Oleh karena itu, untuk memfokuskan pembahasan, maka dilakukan pembatasan topik dan gramatika kalimat tanya. Topik yang dibahas dalam penelitian kali ini adalah pertanyaan konfirmatif tentang fitur-fitur kendaraan bermotor.

Penyusunan gramatika kalimat

Tabel 1. Daftar predikat

Predikat Utama	Ekuivalensi
Diproduksi	Dibuat dikeluarkan
Memproduksi	Membuat mengeluarkan
Mempunyai	Memiliki Ber- (prefiks) menggunakan
Dimiliki	Dipunyai Digunakan Terdapat pada/di Ada pada

Tabel 2.

Penanda	Nama Tanda	Keterangan
< >	Tanda kurung sudut	kata yang berada diantara tanda ini menunjukkan bahwa kata itu adalah sebuah variable
()	Tanda kurung lengkung	keberadaan kata yang berada diantara tanda kurung menunjukkan sifat opsional; boleh ada atau tidak ada
[]	Tanda kurung siku-siku	Keberadaan beberapa tanda ini dalam kalimat menunjukkan posisinya yang saling bisa dipertukarkan dan juga bersifat opsional dengan ketentuan minimal terdapat sebuah dalam kalimat

Proses penyusunan gramatika kalimat tanya konfirmatif dalam penelitian ini difokuskan pada pengenalan kalimat sederhana berbentuk SPO (subjek – predikat - objek). Standar tata bahasa yang digunakan mengacu pada Tata Bahasa Baku Bahasa Indonesia (H. Alwi, 2003). Dalam topik pengenalan pertanyaan konfirmatif tentang fitur-fitur kendaraan bermotor, ternyata kombinasi predikat utama (secara makna) tidak terlalu banyak. Maka untuk menyederhanakan proses penyusunan gramatika, dilakukan pemetaan antara predikat utama

dengan setaranya. Daftar predikat utama dan ekuivalensinya ditunjukkan pada Tabel 1.

Beberapa penanda yang perlu diketahui dalam memahami pola kalimat dalam penyusunan grammar kalimat disebutkan dalam Tabel 2.

Dengan memperhatikan penanda-penanda yang telah dijelaskan pada tabel 2, maka dibuatlah gramatika kombinasi kalimat berita dengan beberapa predikat, seperti tampak pada Gambar 1,2,3,4,5.

```

<Kalimat> : <pabrikannya> memproduksi <kendaraan>
            (<ket_wkt>)
<pabrikannya> : (pabrikannya) <nama_pabrikannya>
<ket_wkt> : (pada) tahun <tahun_produksi>
<kendaraan>: (<jenis_kendaraan>) (<nama_pabrikannya>)
            <nama_kendaraan> ((keluarannya) tahun <tahun>)
    
```

Gambar 1. Gramatika kalimat berita dengan predikat: diproduksi

```

<Kalimat> : <kendaraan> diproduksi [oleh <pabrikannya>]
            [<ket_wkt>]
<pabrikannya> : (pabrikannya) <nama_pabrikannya>
<ket_wkt> : (pada) tahun <tahun_produksi>
<kendaraan>: (<jenis_kendaraan>) (<nama_pabrikannya>)
            <nama_kendaraan> ((keluarannya) tahun <tahun>)
    
```

Gambar 2. Gramatika kalimat berita dengan predikat: memproduksi

```

<kalimat> : <kendaraan> mempunyai <properti_kend>
            <nilai_properti>
<kendaraan> : (<jenis_kendaraan>) (<nama_pabrikannya>)
            <nama_kendaraan> ((keluarannya) tahun
            <tahun>)
    
```

Gambar 3. Gramatika kalimat berita dengan predikat: mempunyai

```

<kalimat> : <kendaraan> mempunyai <properti_kend>
            <nilai_properti>
<kendaraan>: (<jenis_kendaraan>) (<nama_pabrikannya>)
            <nama_kendaraan> ((keluarannya) tahun <tahun>)
    
```

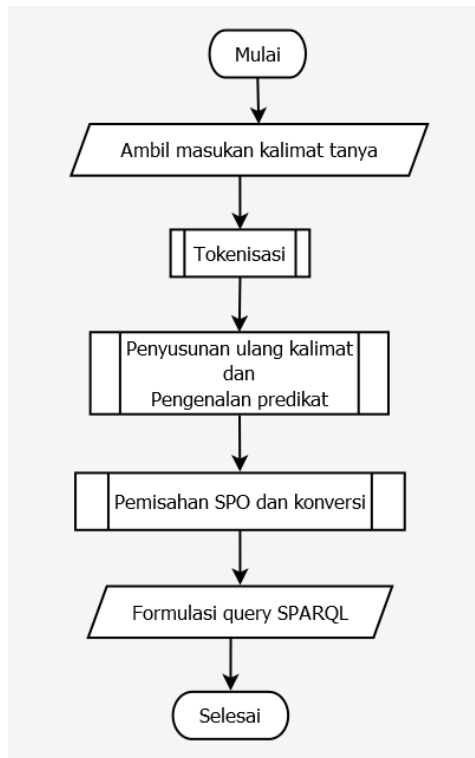
Gambar 4. Gramatika kalimat berita dengan predikat: dimiliki

Untuk mengubah kalimat berita yang dijelaskan grammatikanya pada Gambar 1 menjadi kalimat tanya adalah dengan menambahkan kata tanya *Apakah* di awal kalimat, atau menambah kata *kan* atau *bukan* di akhir kalimat.

Parsing dan Konversi

Pemrosesan bahasa alami (*parsing*) dalam penelitian ini mempunyai tujuan untuk mengecek grammar kalimat input, sedangkan

konversi bertujuan untuk menguraikan kalimat tanya menjadi formulasi query SPARQL. Untuk mewujudkan tujuan tersebut, digunakan pendekatan *top-down parsing*. Dalam implementasinya, *parser* bertumpu pada pengenalan predikat kalimat sederhana. Tahap-tahapan proses yang digunakan dalam pendekatan ini digambarkan dalam bentuk *flowchart*, seperti yang ditunjukkan dalam Gambar 5.



Gambar 5. Diagram alir gambaran umum *parsing*

Tokenisasi

Tokenisasi yang dimaksudkan di sini adalah proses pemecahan kalimat tanya menjadi himpunan frasa/kata yang dianggap penting.

Penyusunan ulang kalimat dan pengenalan predikat

Tujuan dari proses ini adalah untuk meminimalkan kombinasi pola kalimat dalam rangka mempermudah proses konversi dan

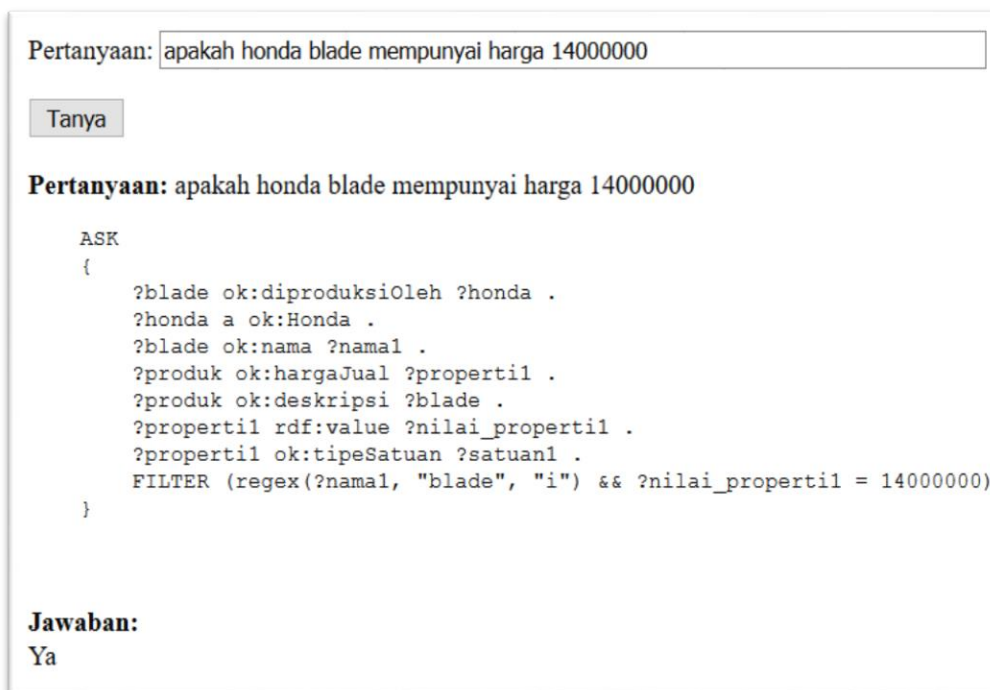
mendata daftar predikat utama yang ada dalam kalimat dan posisinya dalam kalimat tanya input.

Pemisahan SPO dan konversi

Proses ini merupakan proses utama dalam program *parsing*. Tujuannya adalah untuk memetakan kalimat tanya input menjadi pola SPO, kemudian memproses pola SPO tersebut sesuai dengan predikat utamanya. Pemetaan kalimat menjadi himpunan SPO dimaksudkan untuk memudahkan dalam konversi kalimat menjadi formulasi query SPARQL, sebab pernyataan dalam web semantik selalu berupa *triple* (*resource - properti - nilaiProperti*) yang serupa dengan pola SPO.

HASIL DAN PEMBAHASAN

Dalam rangka mengecek keberhasilan penelitian ini, maka perlu dilakukan pengujian. Pengujian dilakukan menggunakan SiTASO (Syarief, 2014) sebagai lingkungan kerja dan OntoMotif (Syarief, 2015) sebagai ontologi utama. Prosedur dalam pengujian ini adalah dengan menginputkan kalimat tanya konfirmatif ke dalam SiTaSO, seperti yang ditunjukkan dalam Gambar 6.



Gambar 6. Contoh pengujian *parsing* dan konversi menggunakan SiTaSO

Berikut akan diberikan beberapa contoh hasil konversi sesuai dengan predikat utamanya:

Jazz R5 diproduksi oleh honda, kan?

```

ASK
{
  ?jazz_r5 ok:nama ?nama1 .
  ?jazz_r5 ok:diproduksiOleh ?honda .
  ?honda ok:nama ?nama_pabrikan1 .
  FILTER (regex(?nama1, "jazz", "i") && regex(?nama1, "r5", "i") && regex(?nama_pabrikan1, "honda", "i"))
}
    
```

honda memproduksi Accord VTI, bukan?

```

ASK
{
  ?accord_vti ok:nama ?nama1 .
  ?accord_vti ok:diproduksiOleh ?honda .
  ?honda ok:nama ?nama_pabrikan1 .
  FILTER (regex(?nama1, "accord", "i") && regex(?nama1, "vti", "i") && regex(?nama_pabrikan1, "honda", "i"))
}
    
```

adakah supra yang menggunakan karburasi injeksi?

```
ASK
{
  ?supra ok:nama ?nama1 .
  ?supra ok:sistemKarburasi ?propertil .
  ?propertil ok:nama ?nilai_propertil .
  FILTER (regex(?nama1, "supra", "i") &&
  regex(?nilai_propertil, "injeksi", "i"))
}
```

apakah honda blade mempunyai harga 14000000

```
ASK
{
  ?blade ok:diproduksiOleh ?honda .
  ?honda a ok:Honda .
  ?blade ok:nama ?nama1 .
  ?produk ok:hargaJual ?propertil .
  ?produk ok:deskripsi ?blade .
  ?propertil rdf:value ?nilai_propertil .
  ?propertil ok:tipeSatuan ?satuan1 .
  FILTER (regex(?nama1, "blade", "i") &&
  ?nilai_propertil = 14000000)
}
```

apakah motor honda supra x keluaran tahun 2015 berkarburasi injeksi

```
ASK
{
  ?supra_x ok:diproduksiOleh ?honda .
  ?honda a ok:Honda .
  ?supra_x a ok:Motor .
  ?supra_x ok:nama ?nama1 .
  ?supra_x ok:diproduksiPadaTahun ?thn1 .
  ?supra_x ok:sistemKarburasi ?propertil .
  ?propertil ok:nama ?nilai_propertil .
  FILTER (regex(?nama1, "supra", "i") &&
  regex(?nama1, "x", "i") && ?thn1 = 2015 &&
  regex(?nilai_propertil, "injeksi", "i"))
}
```

apakah rem depan cakram terdapat pada supra x tahun 2013?

```
ASK
{
  ?supra_x ok:nama ?nama1 .
  ?supra_x ok:diproduksiPadaTahun ?thn1 .
  ?supra_x ok:jenisRemDepan ?propertil .
  ?propertil ok:nama ?nilai_propertil .
  FILTER (regex(?nama1, "supra", "i") && regex(?nama1,
  "x", "i") && ?thn1 = 2013 && regex(?nilai_propertil,
  "cakram", "i"))
}
```

SIMPULAN DAN REKOMENDASI

Dari hasil pengujian, dapat disimpulkan bahwa pendekatan teknik parsing dan konversi dalam penelitian ini berhasil menjawab pertanyaan yang cukup bervariasi berupa kalimat sederhana seputar fitur-fitur kendaraan bermotor. Dalam

penegembangan berikutnya, disarankan untuk melengkapi keterbatasan dalam penelitian ini, yaitu dengan cara menambahkan fitur pengenalan kalimat tanya yang mengandung unsur perbandingan fitur kendaraan dan kalimat tanya kompleks.

DAFTAR PUSTAKA

- Bendi, R. K. (2009). *Sistem Question Answering Sederhana Berbasis Ontologi sebagai Aplikasi Web Semantik*. Yogyakarta: FMIPA UGM.
- Berners-Lee, T. H. (2001). The Semantic Web. *Scientific American*, 284(5), 34–43.
- Cody Kwok, O. E. (2001). Scaling question answering to the web. *ACM Transactions on Information Systems (TOIS)*, 19(3), 242–262.
- H. Alwi, S. D. (2003). *Tata Bahasa Baku Bahasa Indonesia*. Jakarta: Balai Pustaka.
- Rahmad Mahendra, S. D. (2008). Extending an Indonesian Semantic Analysis-based Question Answering System with Linguistic and World Knowledge Axioms. *Proceedings of the 22nd Pacific Asia Conference on Language, Information, and Computation*.
- Septina Dian Larasati, R. M. (2007). Towards a semantic analysis of bahasa Indonesia for question answering. *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*, (pp. 273--280).
- Suryawan, I. W. (2013). *Sistem Question Answering Menggunakan Pendekatan Berbasis Pengetahuan*. Yogyakarta: FMIPA UGM.
- Syarief, M. (2014). *Implementasi Parsing Kalimat Bahasa Indonesia berdasarkan Pengenalan Predikat pada Sistem Tanya Jawab dengan Teknologi Web Semantik*. Yogyakarta: FMIPA UGM.
- Syarief, M. (2015). OntoMotif: Ontologi Pencarian Informasi Kendaraan Bermotor. *Semnastek 2015* (pp. 1-10). Jakarta: Semnastek.