

Ontologi Bahasa Sebagai Basis Pengetahuan untuk Pemrosesan Bahasa Alami

Ontology Bahasa as a Knowledge Base for Natural Language Processing

Fadhila Tangguh Admojo
STMIK Palcomtech, Palembang
e-mail: fadhila.tangguh@gmail.com

Abstrak

Banyak cara yang bisa digunakan oleh manusia untuk dapat berinteraksi dengan manusia lain. Bahasa alami adalah alat komunikasi paling lengkap dan efektif untuk menyampaikan suatu gagasan, pesan, keinginan dan perasaan kepada orang lain. Jika bahasa alami dapat digunakan untuk berinteraksi dengan komputer maka akan sangat memudahkan manusia untuk dapat memberikan berbagai macam tugas kepada komputer. Permasalahan tentang bagaimana komputer memproses dan memahami bahasa alami manusia telah menjadi suatu pembahasan sejak era komputer modern. Sebuah konsep yang dapat digunakan agar komputer mampu memahami bahasa alami manusia yaitu dengan Pemrosesan bahasa alami/Natural Language Processing (NLP). Penelitian NLP untuk menyelesaikan berbagai permasalahan telah banyak dilakukan tetapi yang membahas secara spesifik untuk bahasa Indonesia masih sangat sedikit. Bahasa Indonesia merupakan suatu konsep pengetahuan yang kompleks melibatkan unsur-unsur kebahasaan yang saling terkait. Penelitian ini bertujuan memberikan sebuah gagasan untuk merepresentasikan pengetahuan bahasa Indonesia ke dalam model ontologi.

Kata kunci— Ontologi, Web Semantik, NLP, Pengolahan Bahasa Alami, Basis Pengetahuan

Abstract

There are many ways that humans can use to interact with other humans. Natural language is the most complete and effective communication tool for conveying ideas, messages, desires and feelings to others. If natural language can be used to interact with computers, it will be very easy for humans to be able to provide various tasks to the computer. The problem of how computers process and understand human natural language has been a discussion since the modern computer era. A concept that can be used for computers to be able to understand human natural language is Natural Language Processing (NLP). NLP research to solve various problems has been widely carried out but there is still very few that discusses it specifically for Indonesian. Indonesian is a complex knowledge concept involving interrelated linguistic elements. This research aims to provide an idea for representing Indonesian language knowledge in an ontology model.

Keywords— Ontology, Semantic Web, NLP, Natural Language Processing, Knowledge Base

1. PENDAHULUAN

1.1. Latar Belakang

Banyak cara yang dapat digunakan oleh manusia untuk dapat berinteraksi dengan manusia lain, salah satunya adalah dengan menggunakan bahasa alami/*natural language*. Bahasa alami adalah alat komunikasi yang efektif untuk menyampaikan suatu gagasan, pesan,

dan keinginan kepada orang lain. Sebuah konsep yang dapat digunakan agar komputer mampu memahami bahasa alami manusia yaitu dengan Pemrosesan bahasa alami/*Natural Language Processing* (NLP). [1, 2, 3] menjabarkan komponen NLP merupakan urutan proses yang saling terkait dan membutuhkan basis pengetahuan untuk memproses suatu bahasa. Analisis leksikal membutuhkan pengetahuan tentang kosa kata (leksikon) untuk memahami pembentukan kata. Analisis sintaksis membutuhkan pengetahuan tentang aturan tata bahasa (grammar) untuk memahami struktur suatu kalimat.

Penelitian [4] memaparkan bahwa keterbatasan dalam merepresentasikan pengetahuan menggunakan basis data relasional dapat diatasi dengan menggunakan pemodelan data ontologi. Ontologi dapat menggambarkan bentuk struktur hirarkis yang mengandung definisi kelas, hubungan antar entitas, karakteristik atau propertis dan tata aturan yang berlaku disuatu bidang pengetahuan, artinya dengan menggunakan model ontologi suatu data yang disimpan dapat dipandang sebagai bentuk pengetahuan yang kongkrit karena memiliki arti/makna, saling terkait dan memiliki keterhubungan dengan data yang lain.

Penelitian tentang pemrosesan bahasa alami dengan penggunaan teknologi semantik seperti penelitian [5] yang mengembangkan aplikasi pencarian katalog buku berbasis web. Representasi pengetahuan kepustakaan mengikuti struktur RDF *Simple Knowledge Organization System* (SKOS) standar [6]. Hasil penelitian [5] dapat memproses kalimat perintah sederhana berdasarkan analisis leksikal dan analisis sintaksis menggunakan pola kalimat yang sudah ditentukan.

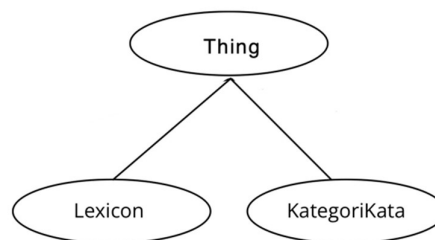
Penggunaan ontology dengan pengetahuan bahasa pernah dibahas oleh [4] yang mengembangkan sistem pencarian informasi berbasis semantik dengan *input* bahasa alami. Hasil penelitian [4] menunjukkan bahwa pencarian informasi dengan penggunaan ontology yang menyimpan fitur-fitur bahasa dapat dilakukan secara lebih dinamis berdasarkan persepsi pengguna baik menggunakan kata kunci atau kalimat sehingga pengguna dapat mencari suatu konten informasi sesuai dengan konteks yang diinginkan.

Penelitian ini berusaha merepresentasikan pengetahuan bahasa ke dalam bentuk ontologi untuk dapat digunakan oleh aplikasi pemrosesan bahasa alami penelitian ini juga berusaha mengkombinasikan ontologi bahasa dengan beberapa domain ontologi lain yang menyimpan konsep pengetahuan yang berbeda.

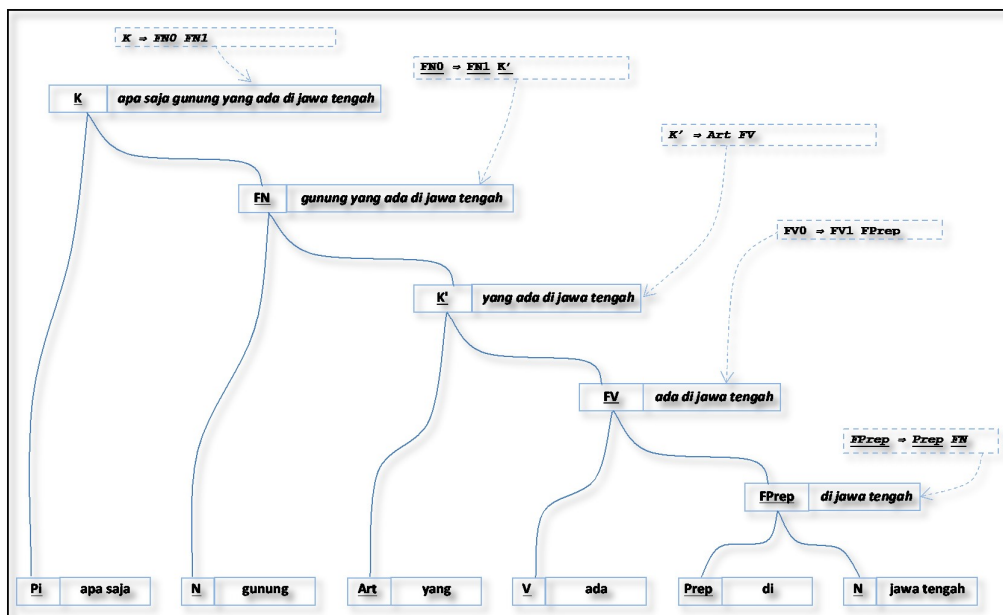
2. METODOLOGI

2.1. Rancangan Ontologi Bahasa

Pengetahuan bahasa yang dibahas dalam penelitian ini berkaitan dengan konsep tentang kata dan hubungan kata. Gambar 1 menjabarkan kelas-kelas yang mewakili konsep kata dan hubungan kata yang dibentuk dalam ontologi yaitu *Lexicon* dan *KategoriKata*. Pengetahuan pada ontologi dideskripsikan ke dalam konsep-konsep yang disusun berdasarkan suatu klasifikasi dan dikelompokkan ke dalam class-class yang sama mengacu pada aturan proses pengurai kalimat yang disajikan pada Gambar 2.



Gambar 1 Kelas pada ontologi



Gambar 2 Ilustrasi pengurai kalimat

Class Lexicon adalah representasi dari konsep kata yang merupakan satuan bahasa terkecil. Setiap individual yang terdapat di dalam *class Lexicon* memiliki kategori, peran dan fungsi di dalam kalimat yang dideskripsikan menggunakan *property word* digunakan untuk mendeskripsikan struktur suatu kata dan *kategori* digunakan untuk mendeskripsikan kategori sintaksis yang dimiliki kata. *Property* yang digunakan untuk mendeskripsikan kata dijabarkan pada Tabel 1.

Tabel 1. *Property* untuk mendeskripsikan konsep kata

Property	Range	Keterangan
<i>word</i>	Xsd:String	deskripsi kata
<i>kategori</i>	KategoriKata	deskripsi kategori sintaksis

Class KategoriKata merupakan representasi dari konsep tentang kategori sintaksis. suatu satuan bahasa dan tidak menggunakan *property* apapun. *Class KategoriKata* disjoint dengan *Class Lexicon*.

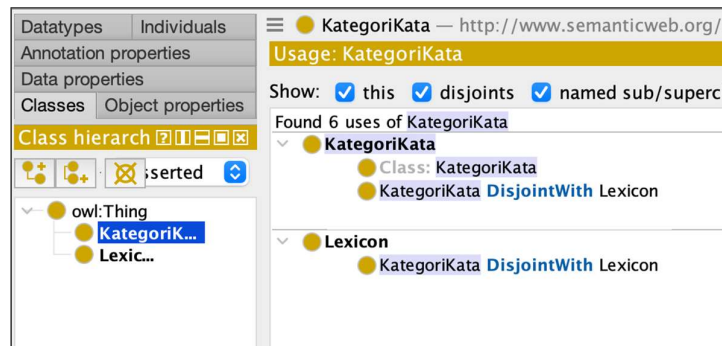
2.2 Implementasi Ontologi Bahasa

Pengetahuan bahasa diimplementasikan ke dalam ontologi menggunakan tool *Protege*. Langkah awal yaitu mendefinisikan *Uniform Resource Identifier* (URI). URI yang digunakan untuk mendefinisikan ontologi bahasa yaitu <http://www.semanticweb.org/ontologies/bahasa-v1>. Deskripsi URI dari dijabarkan pada Gambar 3.



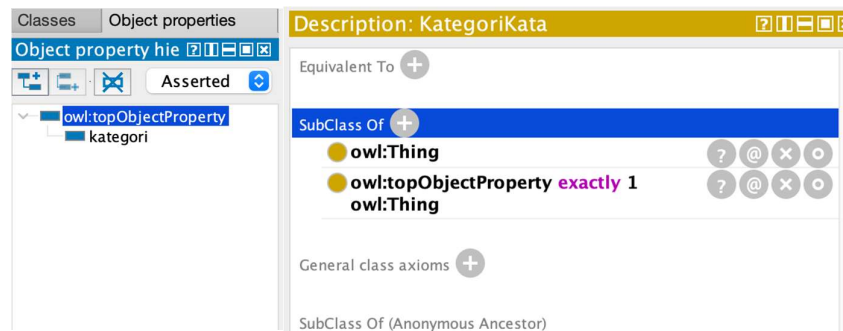
Gambar 3 Implementasi URI pada ontologi Bahasa

Pendefinisian *class* pada *tool Protege* dilakukan menggunakan *tab classes*. *Class* yang didefinisikan ke dalam ontologi bahasa yaitu: *Lexicon* dan *KategoriKata*. Implementasi *class* dan relasi *disjoint* antar *class* dijabarkan pada Gambar 4.



Gambar 4 Implementasi *class* pada ontologi

Tahap selanjutnya dalam implementasi ontologi yaitu mendefinisikan *property*. Pendefinisian *property* pada *tool Protege* dilakukan dengan menggunakan *tab properties*. *Object Property* yang digunakan yaitu *kategori*. Implementasi *property* dan kardinalitas *property* dari *class-class* yang terdapat di dalam ontologi menggunakan *tool Protege* dijabarkan pada Gambar 6.



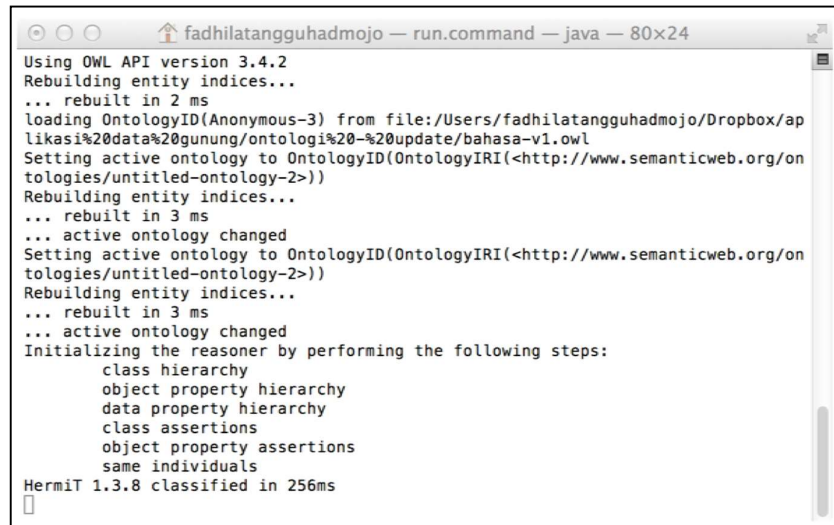
Gambar 6. Implementasi *object property* dan kardinalitas

Tahapan selanjutnya yaitu mendefinisikan *instance* yang merupakan anggota dari *class* yang telah terbentuk. *Instance* didefinisikan menggunakan *tab Individuals* yang terdapat di dalam *tool Protege*. Implementasi dari *instance* untuk *class Lexicon* dan *KategoriKata* *tool Protege* dijabarkan pada Gambar 7.



Gambar 7 Implementasi *instance* dari *class Lexicon*

Langkah terakhir dalam pengembangan ontologi adalah melakukan pemeriksaan inkonsistensi. Pemeriksaan inkonsistensi menggunakan *tool Protege* dilakukan melalui *menu Reasoner → Start reasoner*. Berdasarkan hasil pengecekan pada ontologi bahasa yang disajikan pada Gambar 8 bahwa pada ontologi bahasa tidak terdapat inkonsistensi.

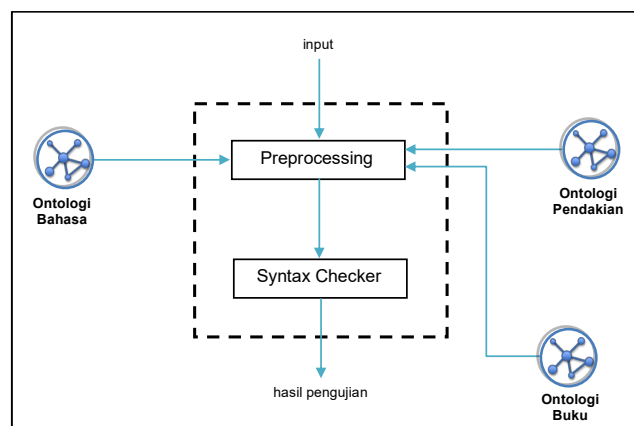


```
Using OWL API version 3.4.2
Rebuilding entity indices...
... rebuilt in 2 ms
loading OntologyID(Anonymous-3) from file:/Users/fadhilatangguhadmojo/Dropbox/ap
likasi%20data%20gunung/ontologi%20-%20update/bahasa-v1.owl
Setting active ontology to OntologyID(OntologyIRI(<http://www.semanticweb.org/on
tologies/untitled-ontology-2>))
Rebuilding entity indices...
... rebuilt in 3 ms
... active ontology changed
Setting active ontology to OntologyID(OntologyIRI(<http://www.semanticweb.org/on
tologies/untitled-ontology-2>))
Rebuilding entity indices...
... rebuilt in 3 ms
... active ontology changed
Initializing the reasoner by performing the following steps:
  class hierarchy
  object property hierarchy
  data property hierarchy
  class assertions
  object property assertions
  same individuals
HermiT 1.3.8 classified in 256ms
```

Gambar 8. Hasil pemeriksaan inkonsistensi pada ontologi bahasa

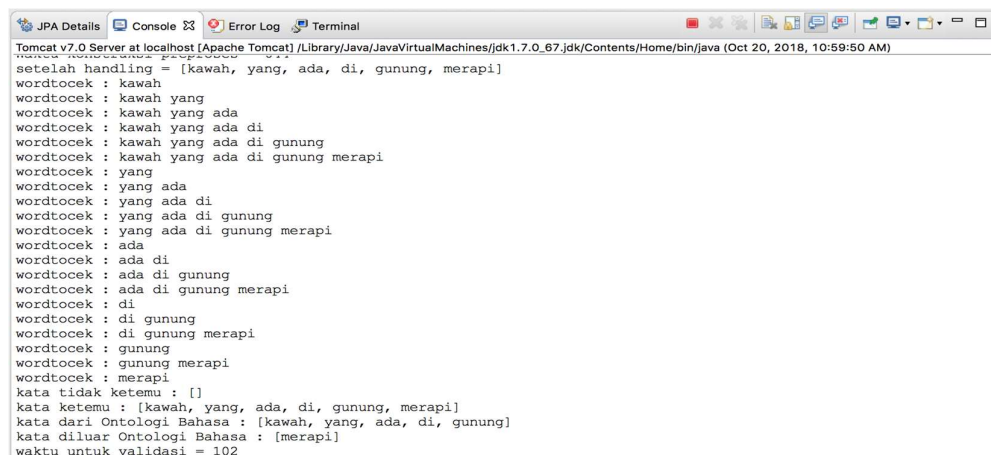
3. HASIL DAN PEMBAHASAN

Pengujian dilakukan untuk mengetahui apakah ontologi bahasa dapat berfungsi sebagai pengetahuan untuk mengolah dan mengurai kalimat. Pengujian terhadap ontologi dilakukan dengan menggunakan aplikasi pemroses bahasa sederhana yang dibangun menggunakan bahasa pemrograman Java dengan Jena Ontology API. Pemroses bahasa terdiri dari dua komponen yaitu *Preprocessing* dan *Syntax Checker*. Komponen *Preprocessing* berfungsi menerima input kemudian melakukan validasi dan memisahkan input menjadi urutan token. Komponen *Syntax Checker* berfungsi melakukan analisis sintaksis dengan mengurai kalimat input menjadi urutan kata membentuk *parse tree*. Arsitektur dari pemroses bahasa yang digunakan untuk pengujian dijabarkan pada Gambar 9.



Gambar 9 Arsitektur pemroses bahasa

Pengujian melibatkan dua ontologi dengan domain pengetahuan yang berbeda, yaitu ontologi pendakian dan ontologi buku. Ontologi pendakian merupakan konsep pengetahuan umum tentang pendakian gunung dan ontologi buku merupakan konsep pengetahuan umum tentang buku. Ontologi pendakian dan ontologi buku dideskripsikan menggunakan konsep SKOS agar struktur skema RDF pada kedua ontologi tetap konsisten meskipun konsep pengetahuan yang direpresentasikan berbeda. Penggunaan domain ontologi lain seperti ontologi pendakian dan buku bertujuan untuk melihat apakah ontologi bahasa bisa diintegrasikan untuk memproses kalimat dengan dua konsep pengetahuan yang berbeda. Misalnya pada kalimat "*apa nama puncak gunung merapi*" ontologi bahasa memiliki pengetahuan tentang kata "*apa, nama, puncak, gunung*" sedangkan kata "*merapi*" merupakan kosakata yang berasal dari ontologi pendakian. Begitu pula dengan kalimat "*siapa pengarang buku head first java*" ontologi bahasa memiliki pengetahuan kata "*siapa, pengarang, buku*" tapi "*head first java*" merupakan kosakata dari ontologi buku. Hasil pengujian untuk melihat penggunaan ontologi dalam melakukan analisis sintaksis disajikan pada Gambar 10 dan gambar 11.

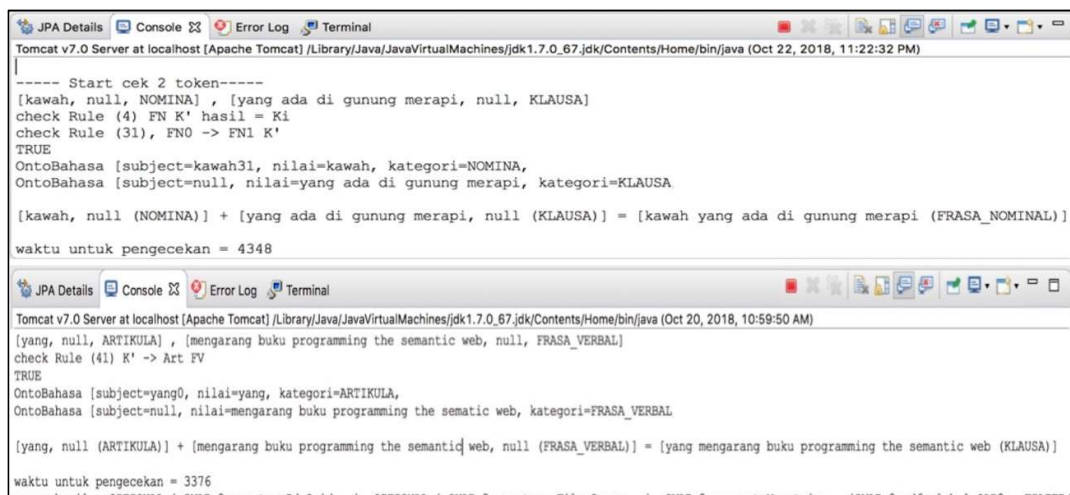


```

Tomcat v7.0 Server at localhost [Apache Tomcat] /Library/Java/JavaVirtualMachines/jdk1.7.0_67.jdk/Contents/Home/bin/java (Oct 20, 2018, 10:59:50 AM)
setelah handling = [kawah, yang, ada, di, gunung, merapi]
wordtocek : kawah
wordtocek : kawah yang
wordtocek : kawah yang ada
wordtocek : kawah yang ada di
wordtocek : kawah yang ada di gunung
wordtocek : kawah yang ada di gunung merapi
wordtocek : yang
wordtocek : yang ada
wordtocek : yang ada di
wordtocek : yang ada di gunung
wordtocek : yang ada di gunung merapi
wordtocek : ada
wordtocek : ada di
wordtocek : ada di gunung
wordtocek : ada di gunung merapi
wordtocek : di
wordtocek : di gunung
wordtocek : di gunung merapi
wordtocek : gunung
wordtocek : gunung merapi
wordtocek : merapi
kata tidak ketemu : []
kata ketemu : [kawah, yang, ada, di, gunung, merapi]
kata dari Ontologi Bahasa : [kawah, yang, ada, di, gunung]
kata diluar Ontologi Bahasa : [merapi]
waktu untuk validasi = 102

```

Gambar 10 Proses pengurai kalimat



```

Tomcat v7.0 Server at localhost [Apache Tomcat] /Library/Java/JavaVirtualMachines/jdk1.7.0_67.jdk/Contents/Home/bin/java (Oct 22, 2018, 11:22:32 PM)
----- Start cek 2 token-----
[kawah, null, NOMINA] , [yang ada di gunung merapi, null, KLAUSA]
check Rule (4) FN K' hasil = Ki
check Rule (31), FN0 -> FN1 K'
TRUE
OntoBahasa [subject=kawah31, nilai=kawah, kategori=NOMINA,
OntoBahasa [subject=null, nilai=yang ada di gunung merapi, kategori=KLAUSA]

[kawah, null (NOMINA)] + [yang ada di gunung merapi, null (KLAUSA)] = [kawah yang ada di gunung merapi (FRASA_NOMINAL)]
waktu untuk pengecekan = 4348

Tomcat v7.0 Server at localhost [Apache Tomcat] /Library/Java/JavaVirtualMachines/jdk1.7.0_67.jdk/Contents/Home/bin/java (Oct 20, 2018, 10:59:50 AM)
[yang, null, ARTIKULA] , [mengarang buku programming the semantic web, null, FRASA_VERBAL]
check Rule (41) K' -> Art FV
TRUE
OntoBahasa [subject=yang0, nilai=yang, kategori=ARTIKULA,
OntoBahasa [subject=null, nilai=mengarang buku programming the semantic web, kategori=FRASA_VERBAL]

[yang, null (ARTIKULA)] + [mengarang buku programming the semantic web, null (FRASA_VERBAL)] = [yang mengarang buku programming the semantic web (KLAUSA)]
waktu untuk pengecekan = 3376

```

Gambar 11 Hasil pengujian menggunakan aturan gramatikal

4. KESIMPULAN

- 1 Ontologi bahasa yang dikembangkan merupakan representasi pengetahuan di bidang linguistik yang meliputi pengetahuan tentang kata dan hubungan kata dalam kalimat dan berdasarkan hasil pengujian ontologi dapat digunakan sebagai dasar pengetahuan untuk memproses kalimat bahasa alami akan tetapi model pengetahuan yang dijabarkan pada penelitian ini masih sangat sederhana dan terbatas, hanya dapat digunakan untuk memproses kalimat sederhana/tunggal dengan menggunakan analisa sintaksis saja.
- 2 Berdasarkan pengujian ontologi bahasa dapat dikombinasikan dengan domain ontologi lain sehingga mampu mengenali dan pemroses kalimat untuk domain yang berbeda namun pengujian yang dilakukan hanya sebatas mengenali kata dan mengurai kalimat menggunakan aturan-aturan. Ketidakmampuan dalam memproses bahasa alami, disebabkan karena kurangnya pengetahuan pada ontologi bahasa, seperti kosakata dan aturan tatabahasa.

DAFTAR PUSTAKA

- [1] Suciadi, J., 2001, Studi Analisis Metode-Metode Parsing dan Interpretasi Semantik pada Natural Language Processing, *Jurnal Informatika*, No. 1, Vol. 2, Mei 2001: 13 - 22.
 - [2] Utami, E. dan Hartati, S., 2007, Pendekatan Metode Rule Based dalam Mengalihbahasakan Teks Bahasa Inggris ke Teks Bahasa Indonesia, *Jurnal Informatika*, N0.1, Vol. 8, 2007.
 - [3] Purwarianti, A., 2009, INAGP: Pengurai Kalimat Bahasa Indonesia sebagai Alat Bantu untuk Pengembangan Aplikasi, *Seminar Nasional Ilmu Komputer dan Aplikasinya (SNIKA)*, No. 1, Vol 4, Bandung, Oktober 2009.
 - [4] F. T. Admojo and E. Winarko, "Sistem Pencarian Informasi Berbasis Ontologi untuk Jalur Pendakian Gunung Menggunakan Query Bahasa Alami dengan Penyajian Peta Interaktif", *IJCCS (Indonesian J. Comput. Cybern. Syst., vol. 10, no. 1, pp. 23-34, 2016.*
 - [5] Andri, 2011, Rancang Bangun Online Public Access Catalog (OPAC) Berbasis Web Sematik, *Tesis*, Program Studi S2 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Gadjah Mada, Yogyakarta.
 - [6] Miles, A. dan Bechhofer, S., 2009, SKOS Simple Knowledge Organization System Reference, Recommendation, World Wide Web Consortium (W3C), <http://www.w3.org/TR/skos-reference>, diakses 20 Oktober 2018.
 - [7] Alwi, H., Dardjowidjojo, S., Lapoliwa, H. dan Moeliono, A. M., 2003, *Tata Bahasa Baku Bahasa Indonesia*, Edisi Ketiga, PT Balai Pustaka (Persero), Jakarta.
-