

1. Pendahuluan

Pada Bab 1 akan dibahas mengenai latar belakang pengambilan topik penelitian, perumusan masalah, tujuan, batasan masalah, metodologi penyelesaian masalah, dan sistematika penulisan

1.1 Latar Belakang

Semantik leksikal adalah cabang semantik yang menyelidiki makna unsur-unsur kosakata suatu bahasa secara umum sebagai satuan mandiri tanpa memandang posisinya dalam kalimat. Sebuah kamus merupakan contoh yang tepat untuk semantik leksikal; makna tiap kata diuraikan di situ (Pateda, 2010). Semantik leksikal berbeda dengan semantik kalimat. Menurut Verhaar (1978: 126), semantik kalimat yang membicarakan hal-hal seperti soal topikalisasi kalimat yang merupakan masalah semantik, namun bukan masalah ketatabahasaan. Mengukur keterkaitan semantik leksikal adalah task penting dalam Pemrosesan Bahasa Alami (PBA).[28] Hal ini seringkali merupakan syarat sebelum melakukan task PBA yang lebih kompleks. Untuk menghasilkan task PBA yang akurat maka digunakan nilai keterkaitan semantik yang akurat.

Ada banyak metode untuk menghitung keterkaitan leksikal semantik, di antaranya: metode path, metode vektor, metode konten informasi, metode gloss, dan metode *hybrid*. Berbagai metode ini tentunya memiliki cara yang unik dalam menghitung nilai keterkaitan semantik. Metode path misalnya. Berdasarkan metode path, keterkaitan antar konsep dapat ditentukan berdasarkan jarak atau panjang path yang menghubungkan konsep di graf semantik berdasarkan tipe edge atau relasi.(Ziqi Zhang, Anna Lisa Gentile, and Fabio Ciravegna 2013) Sedangkan metode vektor berdasarkan Zesch and Guryvech (2010a) adalah metode yang merepresentasikan term atau konsep menggunakan fitur vektor yang didapatkan dari *knowledge base* yang terstruktur. Lain halnya dengan metode konten informasi. Metode konten informasi menggunakan informasi yang sama, yang di taksonomi seringkali ditentukan berdasarkan lcs (longest common subsummer). (Ziqi Zhang, Anna Lisa Gentile, and Fabio Ciravegna 2013) Metode gloss menggunakan gloss dari kata yang akan diteliti yang didapat dari WordNet. Gloss adalah keterangan atau makna dari suatu kata yang terdapat di dalam WordNet. Contohnya pada kata *lawyer* memiliki gloss “*a professional person authorized to practice law; conducts lawsuits or gives legal advice*”.

Nilai standar baku dari keterkaitan antar kata ini untuk sementara menggunakan dataset *gold standard* dari inter-Annotator Agreement (IAA). Keterkaitan semantik berbeda dengan kesamaan semantik. Perbedaan yang mendasar antara keterkaitan semantik dan kesamaan semantik adalah pada perhitungannya. Jika kesamaan semantik perhitungannya menggunakan pohon linguistik seperti pada kata *vehicle* yang memiliki turunan berupa kata *car* dan *bus* yang masih dalam satu pohon. Sedangkan keterkaitan semantik bisa melakukan perhitungan pada kata yang berbeda struktur pohon linguistiknya yaitu seperti pada kata *car* dan *driver*, untuk keterkaitan semantik hal ini bisa dihitung berdasarkan metode yang sudah dijelaskan di paragraf sebelumnya walaupun lebih sulit sedangkan pada kesamaan semantik tidak bisa karena kata-kata tersebut memiliki

struktur pohon yang berbeda. Keterkaitan semantik bisa mengukur relasi apapun dari dua term, sedangkan kesamaan semantik hanya bisa mengukur relasi “is a”. Untuk mengukur kesamaan semantik lebih disarankan untuk menggunakan metode path yang menggunakan struktur pohon pada WordNet dengan menggunakan pendekatan relasi “is a”. Sedangkan untuk keterkaitan semantik metode pengukuran yang disarankan adalah metode gloss karena menggunakan pendekatan yang tidak terikat untuk relasi “is a” saja melainkan semua relasi seperti hipernim, holonim, sinonim, dan hiponim.

Aplikasi dari kesamaan semantik antara lain: *biomedical informatics* untuk membandingkan gen dan protein berdasarkan kesamaan fungsinya (Pesquita, Catia; Faria, Daniel; Falcão, André O.; Lord, Phillip; Couto, Francisco M. 2009), *geoInformatics* (Janowicz, K., Raubal, M. and Kuhn, W. (011), dan *computational linguistics* (Budanitsky, Alexander; Hirst, Graeme 2001). Sedangkan aplikasi keterkaitan semantik seperti untuk *clustering* (Matsuo *et al.* 2006; Bollegala, Matsuo dan Ishizuka 2007), *Word Sense Disambiguation (WSD)* (Leacock and Chodorow 1998; Han and Zhao 2010), *Named Entity Recognition (NER)*, dan *Information Retrieval* (Finkelstein *et al.* 2002). Di bidang biomedis, pengukuran keterkaitan semantik ini merupakan teknik penting untuk mencari relasi fungsional antar gen, dan membangun *lexical resources* (Ye *et al.* 2005; Al-Mubaid and Nguyen 2006; Wu *et al.* 2006; Li *et al.* 2010).

Untuk menghitung keterkaitan antar kata yang bisa menghitung untuk semua relasi yaitu hipernim, hiponim, dan holonim maka diimplementasikan metode berbasis gloss. *Input* dari program yang dibuat adalah berupa pasangan kata beserta nilai *gold standard* yang terdapat pada dataset *wordsim353 semantic relatedness* dan *output* berupa nilai keterkaitan pasangan kata dengan menggunakan metode berbasis *gloss* beserta nilai korelasi yang dihasilkan. Hasil dari implementasi ini kemudian akan dibandingkan dengan nilai dari *gold standard* dari IAA untuk dihitung korelasinya. Pada tugas akhir ini dataset yang digunakan adalah *wordsim353 semantic relatedness* yang berisi 252 pasangan kata beserta nilai *gold standard* masing-masing dan dengan *knowledge base* WordNet. Di tahap pengujian juga akan dicari nilai kesamaan semantik metode Wup, Jcn, dan Lin melalui *website WS4J Demo*[44] kemudian akan dihitung korelasinya dengan nilai *gold standard* dan dibandingkan dengan hasil dari korelasi keterkaitan semantik metode berbasis *gloss*.

1.2 Perumusan Masalah

Mengacu pada latar belakang di atas, berikut adalah perumusan masalah pada tugas akhir ini:

1. Bagaimana mengimplementasikan metode berbasis gloss untuk menghitung keterkaitan semantik antar kata dalam bahasa Inggris?
2. Bagaimana menghitung nilai keterkaitan semantik dalam bahasa Inggris pada suatu pasangan kata dengan menggunakan metode berbasis gloss?
3. Apa pengaruh penambahan *gloss holonym* dalam perhitungan nilai keterkaitan sepasang kata dengan metode berbasis gloss?
4. Berapa nilai korelasi terbaik yang didapat dengan menggunakan dataset *wordsim353 semantic relatedness*?

5. Apa saja parameter yang mempengaruhi nilai keterkaitan semantik?
6. Bagaimana perbandingan hasil korelasi keterkaitan semantik metode berbasis gloss dengan kesamaan semantik metode Wup, Jcn, Lin?

1.3 Tujuan

Berdasarkan rumusan masalah yang telah disampaikan, berikut tujuan dari tugas akhir ini:

1. Mengimplementasikan metode berbasis gloss untuk menghitung keterkaitan semantik antar kata dalam bahasa Inggris.
2. Menghitung nilai keterkaitan semantik dalam bahasa Inggris pada suatu pasangan kata dengan menggunakan metode berbasis gloss.
3. Mengetahui pengaruh penambahan gloss holonym dalam perhitungan nilai keterkaitan sepasang kata dengan metode berbasis gloss.
4. Mendapatkan nilai korelasi terbaik menggunakan dataset *wordsim353 semantic relatedness*.
5. Menganalisis parameter yang mempengaruhi nilai keterkaitan semantik.
6. Menganalisis perbandingan hasil korelasi keterkaitan semantik metode berbasis gloss dengan kesamaan semantik metode Wup, Jcn, Lin.

1.4 Batasan Masalah

Batasan masalah dari tugas akhir ini, antara lain :

1. Sistem yang dibuat hanya diaplikasikan untuk mengukur nilai keterkaitan dari sepasang kata.
2. *Knowledge base* yang digunakan adalah Wordnet.
3. Bahasa yang digunakan dalam pengukuran hanya bahasa Inggris.
4. *Dataset* yang digunakan dalam pengukuran keterkaitan semantik adalah *wordsim353 semantic relatedness* yang terdiri dari 252 pasang kata.
5. *Part of Speech* yang digunakan dalam pencarian setiap pasangan kata adalah *noun*.

1.5 Metodologi Penyelesaian Masalah

Penulis menggunakan metodologi berikut untuk menyelesaikan tugas akhir ini:

1. Studi Literatur
Tahap awal dari penyusunan tugas akhir ini yaitu melakukan studi literatur dari berbagai sumber seperti paper, buku, jurnal, internet, dan lain-lain dalam menggali informasi mengenai keterkaitan semantik leksikal, metode *gloss*, WordNet, serta hal-hal lain yang dapat membantu dalam menyusun Tugas Akhir ini.
2. Pengumpulan Data
Pada tahapan ini, penulis mengumpulkan *dataset* yang akan diuji yaitu *wordsim353 semantic relatedness*.
3. Implementasi

Tahap ini merupakan tahap pengimplementasian metode berbasis gloss untuk mengukur keterkaitan semantik leksikal. Selain itu, pengukuran korelasi untuk melihat ketepatan sistem juga diimplementasikan dalam sistem ini. Pembangunan sistem menggunakan bahasa java.

4. Pengujian

Pada tahap pengujian akan digunakan beberapa parameter sebagai acuan dalam proses pengujian. Parameter ini memiliki hubungan yang erat dengan nilai korelasi yang dihasilkan. Setelah proses pengujian selesai, akan dilakukan analisis untuk mencari tahu penyebab hasil yang telah didapatkan dari proses pengujian.

5. Pembuatan Laporan

Membuat laporan Tugas Akhir berdasarkan hasil penelitian.

1.6 Sistematika Penulisan

Tugas Akhir “**Implementasi dan Analisis Keterkaitan Semantik Leksikal dengan Metode Berbasis Gloss**” disusun dengan menggunakan sistematika sebagai berikut:

1. Pendahuluan

Bab 1 menjelaskan latar belakang masalah yang diangkat, rumusan masalah, tujuan, batasan masalah dan metodologi penyelesaian masalah dari tugas akhir ini.

2. Dasar Teori

Bab 2 menjelaskan teori-teori yang mendukung dan terkait dalam tugas akhir ini.

3. Perancangan Sistem

Bab 3 berisi gambaran umum sistem, *flowchart*, dan *preprocessing*.

4. Pengujian dan Analisis

Bab 4 berisi skenario pengujian, hasil pengujian, dan analisisnya.

5. Kesimpulan dan Saran

Bab 5 berisi kesimpulan hasil dari penelitian tugas akhir yang dilakukan dan saran yang membangun untuk kedepannya.