

Analisis dan Implementasi Keterkaitan Semantik dengan Metode berbasis Vektor

Analysis and Implementation of Semantic Relatedness by Vector-Based Method

Chandra Pangestu¹, Ir. Moch. Arif Bijaksana, M.Tech.²

^{1,2}. Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom Bandung

¹chandrapunkhopepush@student.telkomuniversity.ac.id, ²arifbijaksana@telkomuniversity.ac.id

Abstrak

Keterkaitan semantik adalah mengukur kekuatan koneksi semantik yang ada antara unit tekstual, dalam hal ini antar pasangan kata. keterkaitan antara pasangan kata yang mana memiliki keterkaitan makna. Misalnya ingin menentukan besar nilai keterkaitan semantik antara meja dengan kursi. Untuk mengukur nilai keterkaitan antara pasangan kata tersebut ada beberapa metode yang bisa digunakan yang salah satunya adalah metode yang berbasis vektor. Metode berbasis vektor merupakan salah satu metode yang digunakan untuk mengukur seberapa besar keterkaitan semantik antar pasangan kata. Dengan menggunakan metode vektor ini, sepasang kata akan diubah ke dalam bentuk vektor masing-masing dengan membandingkan beberapa fitur antara kedua kata tersebut. Pada Tugas Akhir ini diimplementasikan dalam bentuk aplikasi untuk mengukur keterkaitan antara sepasang kata dengan metode berbasis vektor, dengan fitur yang didapat dari kedua kata tersebut dengan menggunakan WordNet, metode tf-idf, dan wordsim353 sebagai dataset pengujian yang berisi 252 pasangan kata beserta nilai *gold standard*. Hasil penelitian yang didapat menggunakan metode berbasis vektor pada wordsim353 didapatkan nilai korelasi sebesar 0,42 dengan menggunakan *gloss* dan *meronim*. Dan parameter yang mempengaruhi nilai keterkaitan semantik yang baik adalah dengan menggunakan *gloss* dan *hipernim*.

Kata Kunci: WordNet, vektor, keterkaitan semantik, tf-idf, pasangan kata, dataset.

Abstract

The semantic relatedness to measure the strength of semantic connections that exist between textual units, in this case between pair of words. the relationship between pair of words which have relatedness meaning. For example, want to determine the value of semantic relatedness between table with chair. To measure the value of the relationship between the pair of word there are several methods that can be used, one of which is a vector-based method. Vector-based method is one method used to measure how much the semantic relationship between pair of word. By using this vector, pair of word to be converted into a vector form each by comparing features between the two words. In this final project is implemented in the form of applications to measure the relationship between pair of words with vector-based method, with features that are obtained from both word using WordNet, tf-idf method and wordsim353 as testing dataset containing 252 pairs of words with the goal standard value. Results of the study were obtained using a vector-based method at wordsim353 obtained correlation scored 0,42 by using gloss and meronim. And the parameter that affect the value of a good semantic relatedness is using gloss and hipernim.

Keyword: WordNet, vector-based method, semantic relatedness, corpus, pair of words, dataset.

1. Pendahuluan

Semakin berkembangnya teknologi dewasa ini sangat pesat di berbagai bidang, salah satu bidangnya yaitu teknologi informasi. Dalam perkembangan teknologi informasi ada berbagai cara untuk mendapatkan informasi yang ada, salah satunya adalah mendapatkan informasi seberapa besar keterkaitan antar sepasang kata dalam suatu dokumen.

Keterkaitan antar pasangan kata atau yang lebih dikenal dengan keterkaitan semantik merupakan tugas mencari dan mengukur kekuatan koneksi semantik yang ada antara unit tekstual, dalam hal ini antara pasangan kata. keterkaitan antara sepasang kata yang mana sepasang kata tersebut memiliki keterkaitan makna. Keterkaitan semantik mencakup hubungan yang luas, konsep yang mencakup kesamaan (*sinonim*) serta hubungan lainnya seperti *is-a-kind-of* (*hipernim*, *hiponim*), *is-a-part-of* (*holonim*, *meronim*), dan *is-the-opposite-of* (*antonim*).^[15] Untuk mengukur seberapa besar keterkaitan antara pasangan kata tersebut ada beberapa metode yang bisa digunakan yang salah satunya adalah metode yang berbasis vektor.

Metode berbasis vektor merupakan salah satu metode yang digunakan untuk nilai keterkaitan semantik antar pasangan kata. Dengan menggunakan metode vektor ini, sepasang kata diubah ke dalam bentuk vektor yang masing-masing membandingkan semua fitur antara kedua kata tersebut. Metode vektor umumnya lebih luas daripada yang lain pada sumber informasinya umumnya langsung.[24] juga menunjukkan bahwa metode berbasis vektor konsep dapat dengan mudah diadaptasi di tiga basis yang berbeda pengetahuan (WordNet, Wikipedia, dan Wiktionary) dan mendapatkan hasil yang sebanding. Hal ini menunjukkan bahwa metode berbasis vektor bisa menjadi pilihan yang lebih baik ketika menangani domain baru dan dataset. Dalam metode basis vektor ini salah satu fungsi yang digunakan untuk mengukur keterkaitan antara sepasang kata yaitu fungsi kesamaan kosinus. Fungsi kesamaan kosinus atau *cosine similarity function* adalah fungsi yang pada umumnya digunakan pada sebagian besar metode basis vektor. Karena fungsi tersebut berbeda secara signifikan dalam hal bagaimana vektor fitur dibangun. Fungsi ini umumnya mengukur keterkaitan daripada kesamaan.[25] Untuk membangun keterkaitan semantik dengan metode berbasis vektor digunakan metode tf-idf. Tf-idf adalah indikator kekhususan fitur katanya.[14] Tf-idf digunakan sebagai faktor bobot dalam pencarian informasi dan *text mining*. [22]

Pada studi kasus kali ini akan diimplementasikan dalam bentuk aplikasi untuk mengukur keterkaitan antar pasangan kata dengan metode berbasis vektor, fungsi kesamaan kosinus, metode tf-idf, menentukan fitur yang akan digunakan, menggunakan dataset *wordsim353 semantic relatedness* dan *knowledge base* WordNet. Dengan tujuan untuk menerapkan metode berbasis vektor untuk mengimplementasikan keterkaitan semantik antar sepasang kata dalam bahasa Inggris, melakukan perhitungan nilai terhadap keterkaitan semantik pasangan kata dalam bahasa Inggris yang telah diimplementasikan, mendapatkan nilai korelasi terbaik menggunakan dataset *wordsim353 semantic relatedness*, dan menganalisis parameter yang mempengaruhi nilai keterkaitan semantik sepasang kata.

2. Dasar Teori

2.1 Semantik

Semantik adalah cabang linguistik (ilmu bahasa) yang mempelajari makna yang terkandung pada suatu bahasa, kode, atau jenis representasi lain.

2.2 Keterkaitan Semantik

Keterkaitan Semantik adalah alat matematika yang digunakan untuk memperkirakan kekuatan hubungan semantik antara unit bahasa, konsep atau contoh, melalui deskripsi numerik yang diperoleh sesuai dengan perbandingan informasi pendukung maknanya atau menjelaskan sifat mereka. Keterkaitan semantik diartikan juga bagaimana dalam hal ini sepasang kata memiliki besar nilai keterkaitan tertentu.[6]

2.3 Metode Berbasis Vektor

Metode Berbasis Vektor adalah salah satu metode untuk mendapatkan keterkaitan antar pasangan kata. Dalam hal kategori yang berbeda, ada preferensi yang jelas untuk metode berbasis vektor. Sebagai contoh, semua metode berbasis vektor diperkenalkan sejak tahun 2006, dan itu menjelaskan sebagian besar (hampir 40%) dari semua metode diperkenalkan sejak saat itu. merupakan istilah atau konsep menggunakan fitur vektor yang berasal dari basis pengetahuan terstruktur, daripada menggunakan jumlah yang mendukung suatu kejadian atau konteks. Fitur dari metode vektor banyak digunakan untuk data representasi. Untuk metode berbasis vektor, konsep dan bentuk *lexicalized* mereka bisa diwakili berdasarkan fitur dikodekan dalam basis pengetahuan.

Contoh : *a machine for performing calculations automatically*

proses [a] [machine] [for] [performing] [calculations] [automatically]

vektor v_1 v_j v_k v_l v_m v_n

2.4 Fungsi Kesamaan Kosinus

Fungsi kesamaan kosinus atau *similarity cosine function* adalah fungsi yang digunakan oleh sebagian besar metode berbasis vektor. Namun, berbeda secara signifikan dalam hal bagaimana fitur vektor dibangun. Karena metode tersebut tidak mengeksplorasi sifat hirarkis basis pengetahuan, fungsi tersebut umumnya mengukur keterkaitan daripada kesamaan. [25] Adapun rumus dari fungsi kesamaan kosinus yaitu :

$$\text{cosine}(\vec{v}_1, \vec{v}_2) = \frac{\vec{v}_1 \cdot \vec{v}_2}{\|\vec{v}_1\| \cdot \|\vec{v}_2\|}$$

Gambar 1 rumus kesamaan kosinus.

Dilihat dari rumus pada **Error! Reference source not found.**, dimana v_1 dan v_2 adalah vektor yang kita tentukan,[25] dimana $v_1 \cdot v_2$ adalah fitur yang sama yang dimiliki vektor 1 dan vektor 2, sedangkan $\|v\|$ adalah panjang vektor $\|v\| = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$ yang mana semua vektor pada kata dikuadratkan lalu dijumlahkan,

dan terakhir diakarkan. Jika nilai kosinus yang didapat semakin besar maka keterkaitan antara kedua kata yang dimasukkan adalah besar atau semakin terkait.

2.5 WordNet

WordNet adalah database leksikal untuk bahasa Inggris. Ini kelompok kata bahasa Inggris ke set sinonim yang disebut synsets, memberikan definisi dan contoh penggunaan, dan mencatat sejumlah hubungan di antara set sinonim atau anggotanya. WordNet dapat dipandang sebagai kombinasi dari kamus dan tesaurus. Sementara itu diakses oleh pengguna manusia melalui web browser, penggunaan utamanya adalah dalam analisis teks otomatis dan aplikasi kecerdasan buatan. Database dan perangkat lunak telah dirilis di bawah lisensi BSD gaya dan tersedia secara bebas untuk di-download dari situs WordNet. [6]

2.7. TF-IDF

TF-IDF adalah statistik numerik yang dimaksudkan untuk mencerminkan betapa pentingnya sebuah kata pada dokumen yang ada. Hal ini sering digunakan sebagai faktor bobot dalam pencarian informasi dan pertambahan teks. Nilai TF-IDF meningkat secara proporsional dengan jumlah kali sebuah kata muncul dalam dokumen, tetapi diimbangi dengan frekuensi kata dalam dokumen yang ada, yang membantu untuk menyesuaikan fakta bahwa beberapa kata lebih sering muncul pada umumnya.[22]

2.7.1. Term Frequency (TF)

Term Frequency (TF) menyatakan jumlah banyak keberadaan suatu term atau kata dalam sebuah dokumen, adapun contoh perhitungan tf sebagai berikut :

Contoh : saya dan dia pergi ke rumah saya.

TF saya = 2

2.7.2. Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) berfungsi mengurangi bobot suatu term jika kemunculannya banyak tersebar di seluruh koleksi dokumen kita. Rumusnya adalah dengan inverse document frequency. Document frequency adalah seberapa banyak suatu term muncul di seluruh document yang diselidiki[7].

Contoh : jika kata “saya” muncul dalam 3 dokumen dari 5 dokumen yang ada, maka nilai “saya” adalah

IDF saya = $\log_{10}(5/3)$

IDF saya = 0.477

$$idf_i = \log_{10} (N/df_i)$$

Gambar 2 Rumus IDF

2.8. Korelasi Pearson

Korelasi Pearson merupakan suatu Teknik Statistik yang dipergunakan untuk mengukur kekuatan hubungan 2 Variabel dan juga untuk dapat mengetahui bentuk hubungan antara 2 Variabel tersebut dengan hasil yang sifatnya kuantitatif. Nilai korelasi selalu dalam range $-1 \leq r \leq +1$. [4]

Adapun rumus korelasi pearson adalah, sebagai berikut :

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\{n\sum x^2 - (\sum x)^2\} \{n\sum y^2 - (\sum y)^2\}}}$$

Dimana :

n = Banyaknya Pasangan data X dan Y

$\sum x$ = Total Jumlah dari Variabel X

$\sum y$ = Total Jumlah dari Variabel Y

$\sum x^2$ = Kuadrat dari Total Jumlah Variabel X

$\sum y^2$ = Kuadrat dari Total Jumlah Variabel Y

$\sum xy$ = Hasil Perkalian dari Total Jumlah Variabel X dan Variabel Y

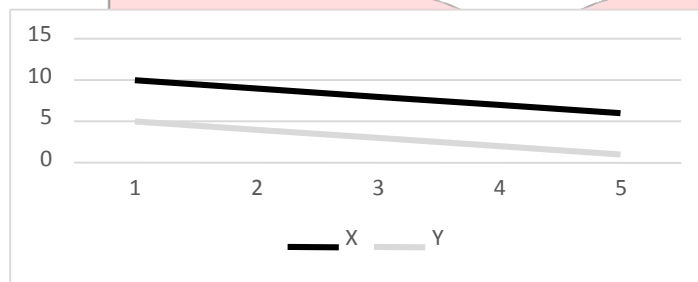
Gambar 3 Rumus korelasi Pearson

Adapun kriteria hubungan, sebagai berikut :

r	Kriteria Hubungan
0	Tidak ada Korelasi
0 – 0.5	Korelasi Lemah
0.5 – 0.8	Korelasi sedang
0.8 – 1	Korelasi Kuat / erat
1	Korelasi Sempurna

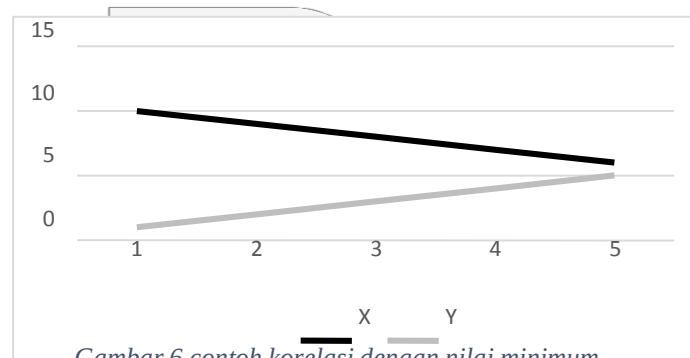
Gambar 4 Kriteria hubungan

Contoh :



Gambar 5 contoh korelasi dengan nilai maksimum

Dapat dilihat gambar di atas adalah contoh grafik korelasi dengan nilai maksimum. Terlihat grafik dengan garis y sejajar atau satu pola terhadap garis x.



Gambar 6 contoh korelasi dengan nilai minimum

Dapat dilihat gambar di atas adalah contoh grafik korelasi dengan nilai maksimum. Terlihat grafik dengan garis y berlawanan terhadap garis x.

2.9. Wordsim 353

WordSim 353 adalah kumpulan tes untuk mengukur kesamaan kata atau keterkaitan, dikembangkan dan dikelola oleh E. Gabrilovich. Dataset ini terdapat 252 pasangan kata dengan nilai gold standardnya.[1] gold standard ditetapkan oleh 13 sampai 16 orang untuk menjadi nilai tetapan pada dataset wordsim353.[26] Dapat dilihat seperti sebagai berikut :

Word 1	Word 2	Gold Standard
computer	keyboard	7.62
Jerusalem	Israel	8.46
planet	galaxy	8.11
canyon	landscape	7.53
OPEC	country	5.63
day	summer	3.94
day	dawn	7.53
country	citizen	7.31
planet	people	5.75
environment	ecology	8.81
Maradona	football	8.62
OPEC	oil	8.59
money	bank	8.5
computer	software	8.5
law	lawyer	8.38
weather	forecast	8.34
network	hardware	8.31
nature	environment	8.31

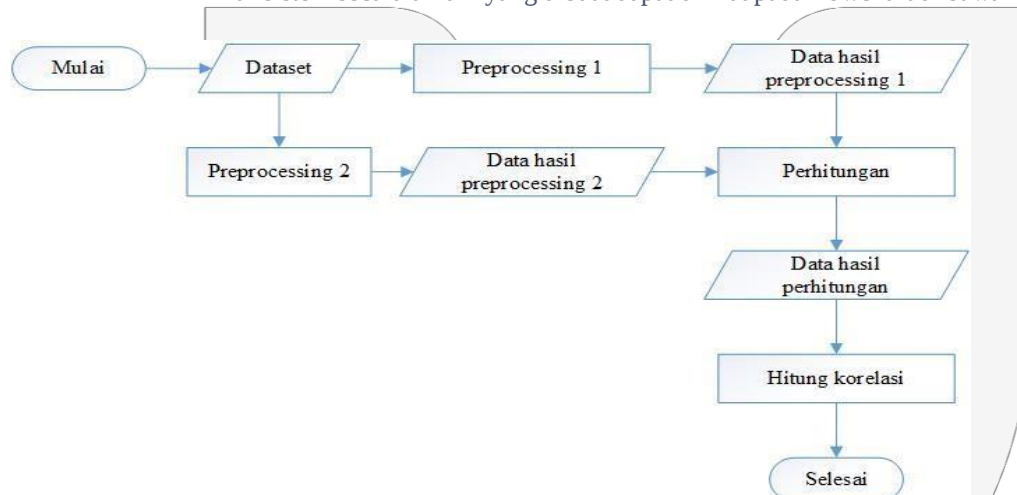
Gambar 7 Potongan Dataset Wordsim 353 untuk keterkaitan kata

2.10 Gambaran Umum Sistem

Sistem untuk membandingkan sepasang kata. Untuk mendapatkan nilai keterkaitannya, ditampilkan outputnya, dan dihitung nilai korelasinya. Hasil nilai keterkaitannya menggunakan metode basis vektor dengan fungsi kesamaan kosinus.

Aplikasi yang akan dibangun ini akan menampilkan input yaitu untuk menginputkan dataset *wordsim353* dimana terdapat 252 pasangan kata dan nilai *gold standard* menjadi tiga kolom yakni kolom kata pertama, kolom kata kedua, kolom nilai *gold standard* dan terakhir ada tombol untuk submit untuk di proses dalam sistem. Nanti menampilkan output yaitu berupa informasi nilai keterkaitan antara sepasang kata, lalu setelah semua pasangan kata dihitung dilanjutkan dengan menghitung nilai korelasi untuk mengetahui nilai yang didapat baik itu lemah, sedang, atau kuat.

Alur sistem secara umum yang dibuat dapat dilihat pada flowchart di bawah.



Gambar 8 Flowchart perancangan sistem

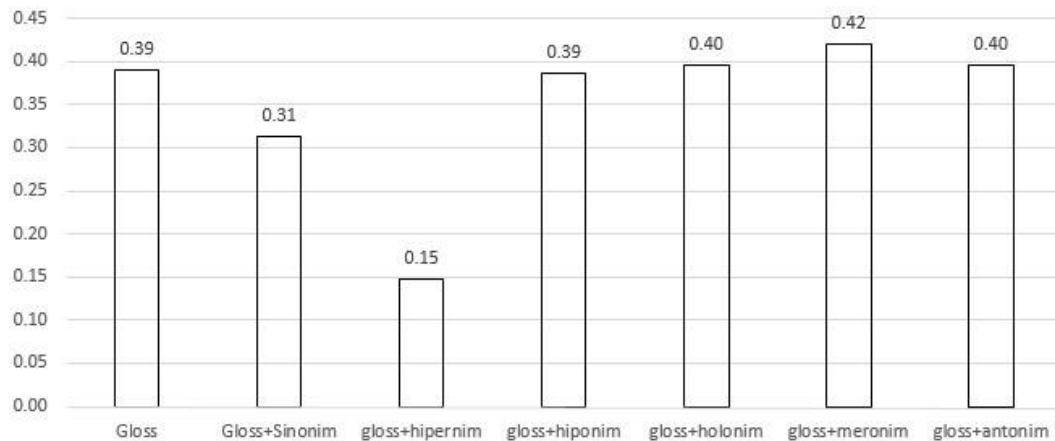
Dari , gambaran umum sistem sebagai berikut :

1. Dataset *wordsim353* berupa file txt yang berisi kata pertama, kata kedua, dan nilai *gold standard* sebanyak 252 pasangan kata dan 252 nilai *gold standard*, yang kemudian semua kata tersebut digunakan untuk memanggil, *gloss* dari sinonim, antonim, hipernim, hiponim, holonim, dan meronim. Yang dijadikan nilai *inverse document frequency* (idf).
2. Kemudian dataset *wordsim353* dipanggil kembali, kali ini setiap sepasang kata dengan nilai *gold standard* untuk memanggil *gloss*, sinonim, antonim, hipernim, hiponim, holonim, dan meronim. Kemudian di *preprocessing* untuk mendapatkan nilai *term frequency* (tf) dari masing-masing fitur pada kata.
3. Setelah itu, nilai tf dari masing-masing fitur pada kata dikalikan dengan fitur yang sama pada fitur dengan nilai idf. Jika semua sudah dikalikan maka setiap fitur sekarang menggunakan nilai tf-idf. Kemudian setiap pasangan kata dihitung dengan fungsi kesamaan kosinus, menggunakan fitur-fiturnya.
4. Jika nilai hasil perhitungan keterkaitan semantik telah didapat sampai pasangan yang paling akhir, kemudian dihitung nilai korelasi menggunakan *pearson's correlation*.

3. Pembahasan

3.1 Analisis Nilai Keterkaitan Pasang Kata Berdasarkan Nilai Korelasi Terbaik

Dilihat dari tujuh hasil korelasi yang didapat, hasilnya ditampilkan dengan grafik pada **Error! Reference source not found.** :



Gambar 9 grafik perhitungan korelasi

Kemudian semua hasil perhitungan korelasi disimpan pada **Error! Reference source not found.** :

Tabel 1 hasil perhitungan korelasi

Unit Linguistik	Nilai Korelasi
Gloss	0.39
Gloss+Sinonim	0.31
gloss+hipernim	0.15
gloss+hiponim	0.39
gloss+holonim	0.40
gloss+meronim	0.42
gloss+antonim	0.40

Dapat dilihat pada **Error! Reference source not found.**, dari tujuh unit linguistik yang digunakan untuk mendapatkan nilai korelasi terbaik untuk keterkaitan semantik, didapatkan nilai korelasi terbaik yaitu 0,42, dengan menggunakan *gloss* dan pengaruh dari penggunaan meronim.

3.2. Analisis Parameter yang Mempengaruhi Nilai Keterkaitan Semantik Terbaik

Untuk menganalisis parameter yang mempengaruhi nilai keterkaitan semantik terbaik, yang mana nilai yang terbaik adalah nilai yang paling dekat dengan nilai *gold standard* karena *gold standard* adalah nilai tetapan untuk setiap pasangan kata pada dataset *wordsim353*, maka digunakan nilai rata-rata selisih setiap nilai *gold standard* dengan nilai dari sistem untuk setiap perhitungan unit linguistik.

Tabel 2 hasil perhitungan nilai rata-rata selisih *gold standard* dengan nilai dari sistem

Unit Linguistik	Nilai Rata-Rata Selisih Nilai Gold Standard dengan Nilai dari Sistem
Gloss	0.501
Gloss+Sinonim	0.491
Gloss+Hipernim	0.465
Gloss+Hiponim	0.493
Gloss+Holonim	0.499
Gloss+Meronim	0.500
Gloss+Antonim	0.500

Dari Tabel 2 hasil perhitungan nilai rata-rata selisih gold standard dengan nilai dari sistem dapat dilihat nilai rata-rata selisih antara nilai *gold standard* dengan nilai dari sistem. Didapatkan nilai rata-rata dengan selisih terkecil adalah menggunakan hipernim. Maka parameter yang mempengaruhi nilai keterkaitan semantik terbaik adalah menggunakan *gloss* dan pengaruh hipernim dari kata yang dibandingkan.

4. Kesimpulan

Dari penelitian perhitungan nilai keterkaitan semantik dengan menggunakan metode berbasis vektor yang telah dilakukan, dapat diambil beberapa kesimpulan sebagai berikut :

1. Dapat mengimplementasikan perhitungan nilai keterkaitan semantik antar pasangan kata menggunakan metode berbasis vektor dengan dataset *wordsim353*, serta menggunakan tf-idf. Maka didapatkan nilai dari perhitungan korelasi pearson yang terbaik yaitu 0,42 dengan menggunakan fitur dari *gloss* dan pengaruh meronim kata yang dibandingkan.
2. Parameter yang mempengaruhi nilai keterkaitan semantik terbaik adalah dengan menggunakan *gloss* dan pengaruh hipernim dari kata yang dibandingkan, dalam perhitungan keterkaitan semantik.

Daftar Pustaka

- [1] Alfonseca.org. “*wordsim353*”, (Online), (<http://alfonseca.org/eng/research/wordsim353.html>. diakses 4 September 2015).
- [2] Budanitsky, A., and Hirst, G. 2006. Evaluating WordNet-based measures of lexical semantic relatedness. *Journal of Computational Linguistics* 32(1), 13–47.
- [3] Eidoon, Zahra; Yazdani, Nasser; Oroumchian, Farhad. 2005. *A Vector Based Method of Ontology Matching*.
- [4] Elektronika, Teknik. “Pengertian Analisis Korelasi Sederhana Rumus Pearson”, (Online), (<http://teknikelektronika.com/pengertian-analisis-korelasi-sederhana-rumus-pearson>. diakses 10 September 2105).
- [5] “Examples.java”, (Online), (http://www.cs.cmu.edu/afs/cs.cmu.edu/project/cmt-40/Nice/Urdu-MT/code/Tools/POS/postagger/mallet_0.4/src/edu/umass/cs/mallet/base/util/resources/wn/Examples.java. diakses 20 September 2015).
- [6] Gabrilovich, Evgeniy, and Markovich, Shaul. 2009. *Wikipedia-based Semantic Interpretation for Natural Language Processing*. Department of Computer Science Technion—Israel Institute of Technology Technion City, 32000 Haifa, Israel.
- [7] Handayanto, Rahmadya Trias. “*Term Frequency dan Invers Document Frequency Tf Idf*”. 25 September 2014. <http://rahmadya.com/2014/09/25/term-frequency-dan-invers-document-frequency-tf-idf>.
- [8] Hassan, Samer and Mihalcea, Rada. 2011. *Semantic Relatedness Using Salient Semantic Analysis*. University of North Texas Denton, Texas.
- [9] Hope, D. 2008. “*Java WordNet::Similarity (beta)*”, (Online), (<http://www.sussex.ac.uk/Users/drh21/>. diakses 7 September 2015).
- [10] Liu, Ying; Pedersen, Ted; Melton-Meaux, Genevieve; McInnes, Bridget T., Pakhomov, Serguei. 2012. *Semantic Relatedness Study Using Second Order Co-occurrence Vectors Computed from Biomedical Corpora, UMLS and WordNet*.
- [11] Martin. “Java”, (Online), (<http://tartarus.org/martin/PorterStemmer/java.txt>. diakses 15 September 2015).
- [12] ” NounOperations.java “, (Online), (https://kask.eti.pg.gda.pl/redmine/projects/ocs/repository/revisions/135/raw/JWNL/src/org/pg/eti/kask/ont/merge_r/util/NounOperations.java. diakses 3 Oktober 2015).
- [13] Organisasi.org. “Makna katapolisemi hipernimi hipernim dan hiponimi hiponim ilmu bahasa Indonesia“, (Online), (<http://www.organisasi.org/1970/01/makna-kata-polisemi-hipernimi-hipernim-dan-hiponimi-hiponim-ilmu-bahasa-indonesia.html>. diakses 1 Desember 2015).
- [14] Patwardhan, Siddharth, and Pedersen, T. 2006. *Using WordNet-based context vectors to estimate the semantic relatedness of concepts*. Proceedings of the EACL 2006 Workshop Making Sense of Sense – Bringing Computational Linguistics and Psycholinguistics Together, pp. 1–8. Stroudsburg, PA, USA: Association for Computational Linguistics.
- [15] Patwardhan, Siddharth; Banerjee, Satyanjeev; Pedersen, Ted. 2003. *Using measures of semantic relatedness for word sense disambiguation*.
- [16] Pesaranhader, Ahmad; Muthaiyah, Saravanan; Pesaranhader, Ali. 2013. *Improving Gloss Vector Semantic Relatedness Measure by Integrating Pointwise Mutual Information*. [17] Sourceforge. “DecisionMethodsList.java“, (Online), (<http://sourceforge.net/p/toxtree/svn/1759/tree/branches/toxtree-3.0.0-SNAPSHOT/toxtree-core/src/main/java/toxTree/tree/DecisionMethodsList.java?format=raw>. diakses 20 September 2015).
- [18] Stanford.edu. 2008. “Stemming and Lemmatization 1”, (Online), (<http://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html>. diakses 27 September 2015).
- [19] Strube, M., and Ponzetto, S. 2006. *WikiRelate! computing semantic relatedness using Wikipedia*. In Proceedings of the 21st National Conference on Artificial Intelligence (AAAI’06), pp. 1419–24. Palo Alto, CA, USA: AAAI Press.
- [20] Wikipedia. “Semantik”, (Online), (<https://id.wikipedia.org/wiki/Semantik>. diakses 1 Maret 2015).

- [21] Wikipedia. "Sinonim", (Online), (<https://id.wikipedia.org/wiki/Sinonim>. diakses 4 Januari 2016).
- [22] Wikipedia. 2012. "Tf-idf", (Online), (<https://en.wikipedia.org/wiki/Tf-idf>. diakses 10 Oktober 2015).
- [23] Zesch, T., and Gurevych, I. 2010a. *Wisdom of crowds versus wisdom of linguists – measuring the semantic relatedness of words*. Natural Language Engineering 16(1), 25–59.
- [24] Zesch, T., and Gurevych, I. 2010b. *The more the better? Assessing the influence of Wikipedia's growth on semantic relatedness measures*. In Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10). Paris, France: European Language Resources Association (ELRA).
- [25] Zhang, Ziqi; Gentile, Anna Lisa; Ciravegna, Fabio. 2012. *Recent advances in methods of lexical semantic relatedness*.
- [26] Radinsky, Kira; Agichtein, Eugene; Gabrilovich, Evgeniy; Markovitch, Shaul. 2011. "a word at a time", (Online), (http://www.cse.cuhk.edu.hk/irwin.king/_media/presentations/a_word_at_a_time.ppt. Diakses 4 Januari 2016).

