

Pembangunan Synsets untuk WordNet Bahasa Indonesia

Tugas Akhir

diajukan untuk memenuhi salah satu syarat

memperoleh gelar sarjana

dari Program Studi Teknik Informatika

Fakultas Informatika

Universitas Telkom

1301144291

I Putu Prima Ananda



Program Studi Sarjana Teknik Informatika

Fakultas Informatika

Universitas Telkom

Bandung

2018

LEMBAR PENGESAHAN

Pembangunan Synsets untuk WordNet Bahasa Indonesia

Building Synsets for Indonesian WordNet

NIM: 1301144291

I Putu Prima Ananda

Tugas akhir ini telah diterima dan disahkan untuk memenuhi sebagian syarat memperoleh

gelar pada Program Studi Sarjana Teknik Informatika

Fakultas Informatika

Universitas Telkom

Bandung, 17 Mei 2018

Menyetujui

Pembimbing I

Pembimbing II

Dr. Moch. Arif Bijaksana, Ir.M.Tech.

NIP: 03650029

Ibnu Asror, S.T., M.T.

NIP: 06840031

Ketua Program Studi
Sarjana Teknik Informatika,

Said Al Faraby, S.T., M.Sc.

NIP: 15890019

LEMBAR PERNYATAAN

Dengan ini saya, I Putu Prima Ananda, menyatakan sesungguhnya bahwa Tugas Akhir saya dengan judul "**Pembangunan Synsets untuk WordNet Bahasa Indonesia**" beserta dengan seluruh isinya adalah merupakan hasil karya sendiri, dan saya tidak melakukan penjiplakan yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan. Saya siap menanggung resiko/sanksi yang diberikan jika dikemudian hari ditemukan pelanggaran terhadap etika keilmuan dalam buku TA atau jika ada klaim dari pihak lain terhadap keaslian karya.

Bandung, 17 Mei 2018

Yang Menyatakan,

I Putu Prima Ananda

Pembangunan Synsets untuk WordNet Bahasa Indonesia

I Putu Prima Ananda¹, Moch. Arif Bijaksana², Ibnu Asror³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

Jl. Telekomunikasi No.1 Terusan Buah Batu Bandung

¹iputuprimaananda@students.telkomuniversity.ac.id, ²arifbijaksana@telkomuniversity.ac.id,

³iasror@telkomuniversity.ac.id

Abstrak

WordNet merupakan database leksikal yang berisi informasi kata, kelas kata, dan definisi seluruh himpunan yang terdapat dalam suatu bahasa. Satuan terkecil dari WordNet adalah synset atau himpunan sinonim yang seluruh anggotanya memiliki arti atau makna yang sama. Peran synset sangat penting bagi WordNet selain merupakan satuan utama, synset menentukan makna dari himpunan kata, dan semua relasi semantik juga menghubungkan synset. Oleh karena itu, pada penelitian ini pembangunan synset khususnya untuk WordNet Bahasa Indonesia dengan menggunakan metode konsep komutatif. Setiap anggota synset dapat saling menggantikan, dimana bila terdapat kata $w1$ yang memiliki sinonim $w2$, dengan menggunakan konsep komutatif maka kata $w2$ harus memiliki sinonim $w1$. Penggunaan metode hukum komutatif dalam pembangunan synset terbukti efektif dengan menghasilkan nilai *precision*, *recall*, dan akurasi F1 sebesar 100%.

Kata kunci : WordNet, synset, sinonim set.

Abstract

WordNet is a lexical database that contains word information, word classes, and definition of all sets contained in a language. The smallest unit of WordNet is a synset or set of synonyms that all member have the same meaning or significance. The role is very important for the WordNet synset. In addition to the main unit, synset determine the meaning of the set words, and all the semantic relationships also connect to synset. Therefore, in this research the building synset especially for WordNet Bahasa by using method of commutative concept. Each synset member can interchanged, if there is a word $w1$ has a synonym $w2$, using concept of commutative then word $w2$ must have a synonym $w1$. Use of the commutative law in the construction method proved effective synset to generate value precision, recall, and accuracy of F1 for 100%.

Keywords: WordNet, synset, synonym set.

1. Pendahuluan

Latar Belakang

WordNet atau yang dikenal dengan Princeton WordNet[11], merupakan sebuah database leksikal bahasa Inggris yang dikembangkan oleh Princeton University, dan dianggap sebagai kamus elektronik terbesar. Princeton WordNet atau PWN dikembangkan oleh ahli leksikographer yang hasilnya dibuat menjadi database leksikal. PWN dibuat secara manual dengan membutuhkan banyak sumber daya seperti ahli bahasa dan waktu sehingga mendapatkan akurasi dan kualitas tinggi [11]. Princeton WordNet berisi informasi tentang 155.000 kata benda (*nouns*), kata kerja (*verbs*), kata sifat (*adjectives*), dan kata keterangan (*adverbs*), kata-kata tersebut dikelompokkan berdasarkan maknanya ke dalam *synonym sets* atau *synsets* [3][10].

Pada perkembangannya, WordNet telah dibuat dalam banyak bahasa lainnya, contohnya adalah Persia WordNet[9] dan Korea WordNet[7] yang dimana dalam proses pembangunannya dilakukan secara otomatis atau menggunakan sumber leksikal yang tersedia. Proses pembangunan WordNet yang pertama dilakukan adalah menghasilkan *synonym sets* atau *synsets*. *Synsets* diperlukan lebih dahulu karena *synsets* merupakan konsep dasar yang mendukung banyak hubungan semantik lainnya di database leksikal. Selain itu, sebuah *synsets* berisi definisi singkat (*gloss*), dimana berisi kalimat yang menggambarkan penggunaan *synsets*[3]. Oleh karena itu, WordNet terutama *synsets* banyak dimanfaatkan dalam berbagai penelitian seperti *computational linguistics*, *natural language processing* (NLP), *information retrieval system*, *text mining*, dan yang lainnya.

Dalam penelitian tugas akhir ini, penulis membangun WordNet Bahasa Indonesia, terbatas hanya pada ekstraksi *synonym sets*, dengan menggunakan *monolingual resources* yang tersedia yaitu Tesaurus Bahasa Indonesia[4].

Tesaurus Bahasa Indonesia merupakan sumber daya yang secara gratis dapat diunduh dan disediakan oleh Pusat Bahasa Indonesia, yang merupakan bagian dari Kementerian Pendidikan Indonesia.

Topik dan Batasannya

Synonym set atau *synset* merupakan sebuah himpunan yang tersusun dari satu atau lebih kata yang memiliki hubungan kesamaan arti atau sinonim[4]. Masing-masing anggotanya dapat saling menggantikan dalam sebagian besar penggunaan kata tersebut dalam sebuah konteks tanpa mengubah sense atau makna kalimat yang memuatnya[4]. Peran synset sangat penting dalam WordNet bahasa apapun, selain karena merupakan satuan utama pada WordNet, synset menentukan makna dari himpunan kata, dan semua relasi semantik juga menghubungkan synset[4].

Pada penelitian[11], pembangunan *synsets* untuk bahasa Indonesia dengan menggunakan *monolingual lexical resources*. Ekstraksi *synsets* dilakukan dengan menggunakan Kamus Besar Bahasa Indonesia[6] dan Tesaurus Bahasa Indonesia[1]. Kedua kamus tersebut dipakai karena merupakan kamus resmi yang telah dipandang baik dan dari penyusun yang sama, Pusat Bahasa Departemen Pendidikan Nasional. Hasil synset didapatkan dengan menggabungkan dua *resources* tersebut. Penelitian[10] menggunakan konsep pemetaan yang didapat dari PWN dan definisi-definisi dari Kamus Besar Bahasa Indonesia (KBBI).

Berdasarkan permasalahan yang telah dijelaskan, rumusan masalah yang dapat diangkat dari penelitian ini adalah sebagai berikut:

1. Bagaimana pembangunan *synsets* Bahasa Indonesia dengan menggunakan metode konsep komutatif?

Batasan-batasan masalah pada penelitian ini guna menyesuaikan kebutuhan dan kemampuan penulis adalah sebagai berikut:

1. *Monolingual resources* yang digunakan pada penelitian ini adalah Tesaurus Bahasa Indonesia
2. Dataset yang digunakan pada penelitian ini diambil secara manual dari tesaurus
3. Fokus utama penelitian ini adalah ekstraksi *synonym sets* dari tesaurus
4. Penulis membuat dataset synset yang telah dihasilkan dari tiap kata secara manual. Hal ini dikarenakan sulitnya mendapatkan data synset dan tidak didistribusikan oleh pemilik atau penulis tidak dapat menemukan dataset synset.

Tujuan

Berdasarkan perumusan masalah dan batasan-batasan yang telah dirumuskan, diharapkan penelitian tugas akhir ini dapat mencapai tujuan penulis, yaitu:

1. Untuk pembangunan *synsets* Bahasa Indonesia dengan menggunakan metode konsep komutatif.

Organisasi Tulisan

Laporan penelitian tugas akhir ini disusun dengan sistematika sebagai berikut:

1. Studi Terkait
Pada bagian ini berisi teori-teori dan literatur terkait untuk mendukung pengerjaan penelitian tugas akhir.
2. Sistem Yang Dibangun
Pada bagian ini berisi tentang proses rancangan dan sistem atau produk yang dihasilkan dari penelitian tugas akhir.
3. Evaluasi
Pada bagian ini berisi tentang hasil pengujian dan analisis hasil pengujian penelitian tugas akhir.
4. Kesimpulan
Pada bagian ini berisikan kesimpulan beserta saran dari hasil penelitian tugas akhir.

2. Studi Terkait

2.1 WordNet

WordNet merupakan sebuah database kamus bahasa Inggris yang dikembangkan oleh Princeton University. Perbedaan WordNet dengan kamus bahasa pada umumnya adalah kamus bahasa memfokuskan pada kata sedangkan WordNet memfokuskan diri pada makna kata[4]. WordNet terdiri dari semua kata yang memiliki arti kata yang

sama dimana setiap synset pada WordNet saling berhubungan. Pada WordNet teridiri dari 4 kelas kata, yaitu kata benda, kata kerja, kata sifat, dan kata keterangan.

WordNet sebagai database leksikal telah dikembangkan dalam lebih dari 70 bahasa lainnya, diantaranya Bahasa Rusia, Spanyol, Turki, Jepang, Thailand, Malaysia[4]. Untuk WordNet dalam Bahasa Indonesia, salah satunya telah dikembangkan oleh Universitas Indonesia[10]. Pada WordNet Bahasa Indonesia diketahui mempunyai 1203 synset dan 1659 kata unik di dalamnya dan jumlah relasi semantik yang dapat dibuat dari synset yang ada mencapai 2261 relasi.

2.2 Synonym Set

Synonym set atau synset merupakan satuan utama yang digunakan dalam WordNet. Synset adalah kumpulan dari satu atau lebih kata yang memiliki makna yang sama atau sinonim[5]. Setiap anggota pada synset dapat menggantikan penggunaan kata tersebut. Sebagai contoh, kata 'bisa' dan 'racun' dapat saling menggantikan dalam konteks semua jenis toksin pada beberapa kelompok hewan, tetapi kata 'bisa' juga dapat diartikan dalam konteks lain, seperti 'mampu' dimana konteks dalam melakukan sesuatu. Setiap kata yang memiliki sinonim atau dapat menggantikan kata lainnya dalam konteks tertentu tidaklah mungkin berasal dari kelas kata yang berbeda[4]. Hal ini dapat terjadi karena sebuah kata dapat termasuk kedalam lebih dari satu kelas kata yang berbeda. Contohnya adalah kata 'satu' dalam Tesaurus Bahasa Indonesia merupakan kelas *noun* dan numerik.

Jika sebuah kata yang tidak memiliki sinonim, maka kata tersebut dianggap sebagai synset dengan anggota tunggal[8]. Hubungan setiap kata dalam synset bersifat komutatif atau simetris, dimana terdapat kata w_1 sinonim dengan kata w_2 , maka w_2 pasti sinonim dengan w_1 pada makna yang sama[4]. Sinonim juga diasumsikan diskrit, dimana dalam satu synset yang memiliki dua atau lebih kata di dalamnya harus saling bersinonim.

2.3 Tesaurus Bahasa Indonesia

Tesaurus berasal dari kata *thesauros*, bahasa Yunani, yang bermakna 'khazanah', sehingga sekarang mengalami perkembangan makna, yakni 'buku yang dijadikan sumber informasi'[1]. Menurut[4], Tesaurus adalah sebuah kamus kumpulan kata yang memiliki arti yang saling terkait. Tesaurus terdiri dari relasi sinonim dan antonim. Terdapat 48.484 item kata Bahasa Indonesia yang terdapat di tesaurus.

Tesaurus dibedakan dari kamus, di dalam kamus dapat dicari informasi tentang makna kata, sedangkan di dalam tesaurus dapat dicari kata yang akan digunakan untuk mengungkapkan gagasan pengguna. Contohnya, bila ingin mencari kata lain untuk kata 'aba-aba', pengguna tesaurus dapat mencarinya pada lema 'aba-aba'.

Tujuan digunakannya tesaurus sebagai dataset adalah tesaurus merupakan kamus besar dimana telah digunakan di beberapa penelitian sebelumnya, selain itu informasi yang terdapat pada tesaurus telah diakui leksigraper dan merupakan sumber daya yang mudah di unduh dan disediakan oleh Pusat Bahasa Indonesia.

2.4 Gold Standard

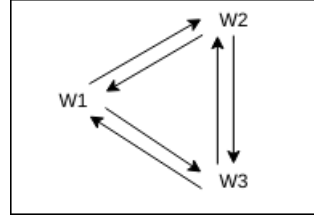
Gold Standard merupakan suatu nilai yang digunakan untuk menganalisis korelasi dari suatu sistem. Standar ini dibuat berdasarkan beberapa pihak ahli pada bidangnya. Pada Tugas akhir ini, dataset yang digunakan sebagai perbandingan dibuat secara manual secara hati-hati sesuai dengan konsep atau hukum komutatif pada synsets. Berikut merupakan isi dari dataset yang dibuat oleh penulis dapat dilihat pada tabel 1.

Tabel 1. Potongan dataset hasil synsets manual

No	Kata	Hasil Synset
1	ahad	ahad, minggu
		ahad, esa, satu, tunggal
2	setanggi	setanggi

2.5 Konsep Komutatif

Digunakannya konsep atau hukum komutatif adalah sesuai dengan sifat synset. Dimana pada satu synset bila terdapat kata w_1 adalah sinonim w_2 , maka kata w_2 pasti sinonim w_1 , hukum ini berlaku untuk semua kata yang terdapat pada sebuah synset.



Gambar 1. Konsep Komutatif

Pada gambar 1 merupakan contoh dari konsep komutatif pada synset. Dimana terdapat synset yang berisi anggota $w1$, $w2$, dan $w3$. Tanda panah $w1$ menuju $w2$ menunjukkan bahwa $w1$ memiliki sinonim $w2$, begitu juga dengan yang lainnya. Gambar 1 menunjukkan sebuah synset harus memiliki keterhubungan komutatif antar semua kata yang ada.

2.6 *F1-score*

Metode yang digunakan dalam mengukur performansi sinonim set pada penelitian tugas akhir ini dengan menggunakan *F1-score*, untuk mengukur performansi menggunakan *F1-score* dibutuhkan *recall* dan *precision*. Recall yaitu mengukur rasio dari jumlah prediksi yang benar terhadap total prediksi yang diharapkan[2]. Formula untuk *recall* yang umum digunakan pada persamaan 1.

$$r = \frac{tp}{tp + fn} \quad (1)$$

$p = precision$.

$tp = true\ positive$, yaitu jumlah data yang diprediksi positif oleh sistem dan dalam kenyataan bernilai positif.

$fn = false\ negative$, yaitu jumlah data yang diprediksi negatif oleh sistem namun dalam kenyataannya bernilai positif.

Precision yaitu mengukur rasio dari jumlah prediksi yang benar terhadap total prediksi[2]. Formula untuk *precision* yang umum digunakan pada persamaan 2.

$$p = \frac{tp}{tp + fp} \quad (2)$$

$p = precision$.

$tp = true\ positive$, yaitu jumlah data yang diprediksi positif oleh sistem dan dalam kenyataan bernilai positif.

$fp = false\ positive$, yaitu jumlah data yang diprediksi positif oleh sistem namun dalam kenyataannya bernilai negatif.

Untuk memudahkan dalam melakukan perhitungan performansi dalam kasus penelitian ini maka formula disederhanakan. Sehingga formula untuk *precision* pada persamaan 3 dan *recall* pada persamaan 4.

$$precision = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{retrieved\ documents\}|} \quad (3)$$

dan,

$$recall = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{relevant\ documents\}|} \quad (4)$$

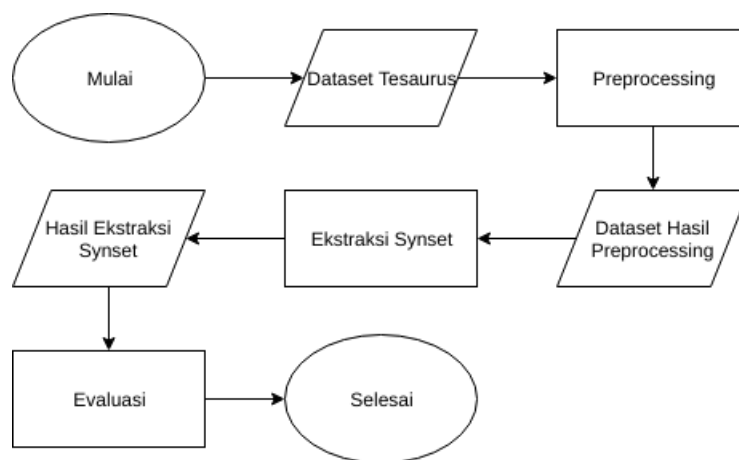
Dimana, *relevant documents* adalah data prediksi, dan *retrieved documents* adalah data dari *gold standard* atau hasil synset yang dibuat secara manual.

Setelah mendapatkan *recall* dan *precision* maka dalam menghitung *F1-score* digunakan formula pada persamaan 5.

$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (5)$$

3. Sistem yang Dibangun

3.1 Rancangan Sistem



Gambar 2. Flowchart gambaran umum sistem

Pada penelitian tugas akhir ini dibangun sebuah sistem yang dapat menghasilkan sinonim set. Metode yang telah diterangkan sebelumnya akan menjadi acuan dalam menentukan pembangunan synset yang sesuai. Gambar (2) menunjukkan alur proses yang menunjukkan alur proses yang dilakukan pada penelitian tugas akhir ini yaitu *Preprocessing*, Ekstraksi Synset, dan Perhitungan Evaluasi. Mula-mula, dataset Tesaurus melalui proses yang pertama yaitu *preprocessing* yang kemudian dihasilkan dataset yang siap untuk diolah, lalu dataset yang dihasilkan dari proses pertama melewati proses yang kedua yaitu Ekstraksi Synset, lalu dataset yang dihasilkan dari proses kedua digunakan untuk Evaluasi yaitu menghitung perfromasi sistem dengan menggunakan *F Measure*.

3.2 Rancangan Data

Dataset yang digunakan pada penelitian tugas akhir ini adalah Tesaurus Bahasa Indonesia yang berisikan kamus Bahasa Indonesia yang memiliki sekelompok kata-kata sinonim. Dataset tersebut menyediakan 48.484 item kata Bahasa Indonesia[4] dan dapat diunduh secara bebas dan disediakan oleh Pusat Bahasa Indonesia. Selanjutnya diambil 30 kata dari Tesaurus secara manual yang akan dicari synsetsnya. Kumpulan kata yang akan menjadi dataset dapat dilihat pada tabel 2.

Tabel 2. kata-kata yang diambil dari tesaurus sebagai dataset

Kata-kata yang diambil dari Tesaurus sebagai Dataset					
1. ahad	6. abu	11. suon	16. fiksi	21. minggu	26. kopiah
2. setanggi	7. peci	12. lilin	17. lamur	22. esa	27. songkok
3. aborsi	8. koran	13. sakat	18. radas	23. pengguguran	28. parafin
4. pekan	9. susur	14. satwa	19. persentase	24. pasar	29. parasit
5. lebu	10. temu	15. binatang	20. bandrek	25. rekan	30. serbat

3.3 Preprocessing

Pada proses pertama pembangunan sistem, dataset tesaurus melalui proses *Preprocessing* yang bertujuan untuk memilih dataset yang akan digunakan dalam proses ekstraksi. Pada proses ini dilakukan pemilahan kata-kata, menghilangkan atribut-atribut yang tidak digunakan seperti *n*, *v*, *adv*, *v*, *cak*, (*cak*), dan yang lainnya. Selanjutnya pada tesaurus yang memiliki sense lebih dari satu dimana pada dataset diketahui dengan penomoran sebelum kata-katanya contohnya kata 'Ahad 1 Minggu; 2 esa, tunggal, satu' akan dikelompokkan sesuai dengan sensenya. Hasil dari *preprocessing* berupa file text dalam format *json* yang telah siap untuk digunakan.

3.4 Ekstraksi Synset

Pada proses ini dilakukan pembangunan sinonim set dari dataset tesaurus dengan menggunakan konsep komutatif yang telah dijelaskan sebelumnya. Sebagai contohnya akan dicari synset 'ahad' sebagai berikut:

ahad 1 Minggu; 2 esa, tunggal, satu

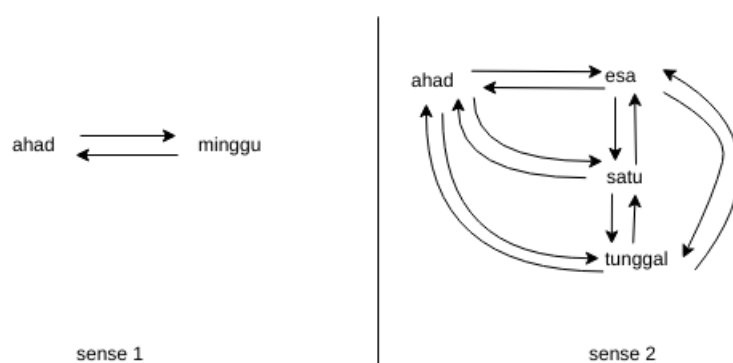
minggu Ahad, pekan

esa ahad, satu, tunggal

tunggal 1 ahad, esa, satu; 2 sendiri, singular, individual, perorangan, solo, 3 satu-satunya, semata wayang, wahid 4 bulat, utuh

satu ahad, eka, esa, homo-, iso-, mono-, se-, suatu, tunggal, uni, unik, wahid

Dari contoh diatas terdapat lema 'ahad' yang memiliki dua sense yaitu sense pertama 'minggu' dan sense kedua 'esa, satu, tunggal'. Selanjutnya akan dicari dan dicocokkan dengan setiap kata pada setiap sense sebelumnya. Pada sense pertama akan dicari pada lema 'minggu' dan juga terdapat kata 'ahad'. Selanjutnya untuk sense kedua lema 'esa' juga terdapat kata 'ahad', lema 'tunggal' terdapat kata ahad pada sense pertama, lema 'satu' terdapat kata 'ahad'. Sehingga sedemikian sesuai dengan konsep komutatif dapat ditampilkan pada gambar (4).



Gambar 3. Hubungan komutatif kata 'ahad'

Pada gambar (4), tanda panah penunjukan hubungan komutatif antar kata. Sehingga synset dari lema 'ahad' terdiri dari 2 synset yaitu 'ahad, minggu', dan 'ahad, esa, satu, tunggal'. Bila terdapat kata yang tidak memiliki hubungan komutatif, maka kata tersebut tidak termasuk anggota synset. Proses inilah yang nantinya akan digunakan pada pembangunan sinonim set.

3.5 Perhitungan Evaluasi

Proses terakhir yang dilakukan adalah perhitungan evaluasi yang berguna untuk mengukur performansi dari sistem yang telah dibuat. Proses evaluasi menggunakan formula *F Measure* yang telah dijelaskan sebelumnya. *Recall* diambil dari nilai data hasil prediksi synset, sedangkan *precision* berasal dari data prediksi sebenarnya.

4. Evaluasi

4.1 Analisis dan Hasil Pengujian

Dalam menguji metode yang digunakan dalam pembangunan synsets pada penelitian tugas akhir ini digunakan dataset hasil synsets secara manual sebagai perbandingan dengan hasil dari program. Berdasarkan hasil pengujian *system* dan hasil evaluasi dengan menggunakan *F Measure* didapatkan hasil yang disertakan didalam tabel 3.

Tabel 3. Tabel Jumlah Synsets yang Dihasilkan dan Hasil Evaluasi

System	Jumlah synset yang dihasilkan (Synsets)	Precision %	Recall %	F1 %
Program	37 Synsets	100%	100%	100%
Manual	37 Synsets			

Sistem yang dibangun dalam penelitian ini menghasilkan 37 synsets dari 30 kata yang dijadikan sebagai dataset. Dari synsets yang dihasilkan diketahui bahwa satu dataset atau kata dapat menghasilkan lebih dari satu synsets. Hasil synsets yang dihasilkan seperti dibawah ini (dalam bentuk file format leksikografer WordNet)[11]:

{ahad, esa, satu, tunggal}
{setanggi}
{minggu, pekan}

Kemudian dari hasil synsets yang didapatkan dari system dilakukan pengujian akurasi dengan menggunakan *F1-score*. Hasil evaluasi yang dilakukan dengan menggunakan *F1 score* menghasilkan nilai *precision* sebesar 100% dan nilai *recall* sebesar 100% sehingga *F1-score* yang didapat sebesar 100%.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil pengujian dan analisis yang telah dilakukan pada penelitian tugas akhir ini, maka dapat ditarik kesimpulan sebagai berikut.

1. Synsets yang dihasilkan terdiri dari synset tunggal atau synset yang memiliki lebih dari satu kata.
2. Setiap kata atau dataset dapat memiliki lebih dari satu .
3. Hasil dari sistem yang telah dibangun dalam penelitian tugas akhir ini dapat menghasilkan nilai akurasi (F1-score) sebesar 100%.
4. Hukum komutatif dapat digunakan untuk membangun synsets untuk WordNet Bahasa Indonesia.

5.2 Saran

Saran untuk penelitian selanjutnya dalam pembangunan synsets WordNet Bahasa Indonesia:

1. Menambahkan dataset lebih banyak baik secara manual atau dibuat dengan program.
2. Mencoba menggunakan *monolingual resources* selain Tesaurus Bahasa Indonesia.

Daftar Pustaka

- [1] E. Endarmoko. *Tesaurus Bahasa Indonesia*. Gramedia Pustaka Utama, 2007.
- [2] J. Euzenat. Semantic precision and recall for ontology alignment evaluation. In *IJCAI*, volume 7, page 348353, 2007.
- [3] C. Fellbaum. *WordNet*. Wiley Online Library, 1998.
- [4] Gunawan. *Akuisisi Gloss Berbasis Ekstraksi Synonym Set Menggunakan Supervised Learning*. Institut Teknologi Sepuluh November, 2016.
- [5] C. Havasi, R. Speer, and J. Alonso. Conceptnet 3: a flexible, multilingual semantic network for common sense knowledge. In *Recent advances in natural language processing*, pages 27–29. Citeseer, 2007.
- [6] T. R. K. B. Indonesia. Kamus bahasa indonesia. *Jakarta: Pusat Bahasa Departemen Pendidikan Nasional*, 2008.
- [7] C. Lee, G. Lee, and S. J. Yun. Automatic wordnet mapping using word sense disambiguation. In *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13*, pages 142–147. Association for Computational Linguistics, 2000.
- [8] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, 3(4):235–244, 1990.

- [9] M. Montazery and H. Faili. Automatic persian wordnet construction. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 846–850. Association for Computational Linguistics, 2010.
- [10] D. D. Putra, A. Arfan, and R. Manurung. Building an indonesian wordnet. In *Proceedings of the 2nd International MALINDO Workshop*, pages 12–13, 2008.
- [11] A. Saputra et al. Building synsets for indonesian wordnet with monolingual lexical resources. In *Asian Language Processing (IALP), 2010 International Conference on*, pages 297–300. IEEE, 2010.

Lampiran

Berikut merupakan lampiran-lampiran pada penelitian tugas akhir ini.

Tabel 4. Tabel hasil synset program

Kata ke-	Kata	Synset Program
1	ahad	[[['ahad', 'minggu']], [['ahad', 'esa', 'satu', 'tunggal']]]
2	setanggi	[[['setanggi']]]
3	aborsi	[[['aborsi', 'pengguguran']]]
4	pekan	[[['pasar', 'pekan', 'rekan']], [['minggu', 'pekan']]]
5	lebu	[[['abu', 'duli', 'lebu']]]
6	abu	[[['abu', 'abuk', 'debu'], ['abu', 'debu', 'duli'], ['abu', 'duli', 'lebu']]]
7	peci	[[['kopiah', 'peci', 'songkok']]]
8	koran	[[['harian', 'koran', 'surat kabar']]]
9	susur	[[['susur']]]
10	temu	[[['jumpa', 'temu']]]
11	suar	[[['pijar', 'suar']], ['suar']]
12	lilin	[[['lilin', 'parafin']]]
13	sakat	[[['benalu', 'parasit', 'sakat']]]
14	satwa	[[['binatang', 'hewan', 'satwa']]]
15	binatang	[[['binatang', 'hewan', 'satwa']]]
16	fiksi	[[['fantasi', 'fiksi']]]
17	lamur	[[['lamur', 'rabun']]]
18	radas	[[['perkakas', 'radas']]]
19	presentase	[[['persentase']]]
20	bandrek	[[['bandrek', 'serbat']]]
21	minggu	[[['ahad', 'minggu'], ['minggu', 'pekan']]]
22	esa	[[['ahad', 'esa', 'satu', 'tunggal']]]
23	pengguguran	[[['aborsi', 'pengguguran']]]
24	pasar	[[['pasar', 'pekan', 'rekan']]]
25	rekan	[[['pasar', 'pekan', 'rekan']]]
26	kopiah	[[['kopiah', 'peci', 'songkok']]]
27	songkok	[[['kopiah', 'peci', 'songkok']]]
28	parafin	[[['lilin', 'parafin']]]
29	parasi	[[['benalu', 'parasit', 'pasilan'], ['benalu', 'parasit', 'sakat']]]
30	serbat	[[['bandrek', 'serbat']]]

```

Jumlah synsets program : 37
Jumlah synsets manual : 37
Jumlah synsets yang sama : 37
precision 1.0
recall 1.0

```

```

akurasi hasil : 100.0

```

Gambar 4. Hasil akurasi program