Pembangunan Synonym Set untuk WordNet Bahasa Indonesia dengan Metode Komutatif

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 Synonym Set untuk WordNet Bahasa Indonesia dengan Metode Komutatif

Building Synonyms Set for Indonesian WordNet with Commutative Method

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, 27 Juni 2018
Menyetujui

Pembimbing I	Pembimbing II
Dr. Moch. Arif Bijaksana,Ir.M.Tech.	Ibnu Asror,S.T.,M.T.
NIP: 03650029	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 Synonym Set untuk WordNet Bahasa Indonesia dengan Metode Komutatif**" beserta dengan seluruh isinya adalah merupakan hasil karya sendiri, dan saya tidak melakukan penjiplakan yang tidak sesuai dengan etika keilmuan yang belaku 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, 27 Juni 2018

Yang Menyatakan,

I Putu Prima Ananda

Pembangunan Synonym Set untuk WordNet Bahasa Indonesia dengan Metode Komutatif

I Putu Prima Ananda[1], Moch. Arif Bijaksana[2], Ibnu Asror[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
Jl. Telekomunikasi No.1 Terusan Buah Batu Bandung
[1]iputuprimaananda@students.telkomuniversity.ac.id, [2]arifbijaksana@telkomuniversity.ac.id,
[3]iasror@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 $w1$. Ide pembangunan synset ini diambil dari penelitian [1], dimana perbedaannya pada penelitian ini pembangunan synset dilakukan dengan menggunakan metode komutatif. Performansi yang dihasilkan dari implementasi komutatif terhadap teori komutatif menghasilkan nilai F1 sebesar 100%**.**

Kata kunci : metode komutatif, synonim set, synset, WordNet.

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 $w1$. This idea was taken from research [1], where the difference in this study focus on using the commutative method. The performance resulting from the commutative implementation of the commutative theory result an F1 for 100%**.**

Keywords: commutative method, synonym set, synset, WordNet.

1. Pendahuluan

Latar Belakang

WordNet atau yang dikenal dengan Princeton WordNet[2], merupakan 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 memiliki kualitas tinggi [2]. Pada penelitian [3], WordNet berisi informasi tentang 155.000 kata benda (*nouns*), kata kerja (*verbs*), kata sifat (*adjectives*), dan kata keterangan (*advebrs*), kata-kata tersebut dikelompokkan berdasarkan maknanya ke dalam synonim set (synset) atau kumpulan sinonim yang memiliki makna sama.

Pada perkembangannya, WordNet telah dibuat dalam bahasa lainnya, contohnya adalah Persian WordNet[4] dan Korean WordNet[5] yang dimana dalam proses pembangunannya dilakukan secara otomatis atau menggunakan sumber leksikal yang tersedia. Struktur WordNet berisi informasi kata, kelas kata, dan definisi dari seluruh himpunan kata yang terdapat pada suatu bahasa. Ketiga elemen tersebut menjadi entitas tunggal yang saling berelasi[1]. Pada kamus bahasa pada umumnya satuan terkecilnya adalah kata, berbeda pada WordNet satuan terkecilnya adalah synset atau himpunan sinonim yang memiliki makna yang sama. Proses pembangunan WordNet yang pertama dilakukan adalah menghasilkan synonim set atau synset. Synset diperlukan lebih dahulu karena

merupakan konsep dasar yang mendukung banyak hubungan semantik lainnya di database leksikal. Selain itu, sebuah synset berisi definisi singkat (*gloss*), dimana berisi kalimat yang mengambarkan penggunaan synset[3]. Oleh karena itu, WordNet terutama synset banyak dimanfaatkan dalam berbagai penelitian seperti *computational linguistics*, *natural language processing* (NLP), *information retrieval system*, *text mining*, dan yang lainnya. Pada setiap synset atau himpunan sinonim memiliki relasi yang komutatif dimana setiap kata pada synset harus saling berhubungan satu sama lain. Sebagai contohnya bila terdapat *word1* yang memiliki sinonim *word2*, maka *word2* pasti memiliki sinonim *word1*. Hal tersebut juga berlaku untuk setiap kata yang terdapat pada himpunan synset. Konsep komutatif dapat digunakan sebagai pembangunan synset dalam berbagai Bahasa.

Dalam penelitian tugas akhir ini, penulis membangun WordNet Bahasa Indonesia, terbatas hanya pada pembangunan synset, dengan menggunakan *monolingual resources* yang tersedia yaitu Tesaurus Bahasa Indonesia[1]. Pada Tesaurus Bahasa Indonesia konsep komutatif tidak selalu terjadi. Pada penelitian ini proses ekstraksi synset Bahasa Indonesia dengan Tesaurus menggunakan metode komutatif pada synset.

Topik dan Batasannya

Synonym set atau synset merupakan sebuah himpunan yang tersusun dari satu atau lebih kata yang memiliki hubungan kesamaan arti atau sinonim[1]. Masing-masing anggota synset dapat saling menggantikan dalam sebagian besar penggunaan kata tersebut dalam sebuah konteks tanpa mengubah sense atau makna kalimat yang memuatnya[1]. Kumpulan kata pada synset tersebut dapat sebagai konsep komutatif. 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[1].

Pada penelitian[2], pembangunan *synsets* untuk bahasa Indonesia dengan menggunakan *monolingual lexical resources*. Ekstraksi *synsets* dilakukan dengan menggunakan Kamus Besar Bahasa Indonesia[6] dan Tesaurus Bahasa Indonesia[7]. 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[8] menggunakan konsep pemetaan yang didapat dari PWN dan definisi-definisi dari Kamus Besar Bahasa Indonesia (KBBI).

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

1. Bagaimana implementasi metode komutatif untuk pembangunan synsets Bahasa Indonesia?
2. Bagaimana performansi algoritma metode komutatif untuk pembangunan synsets Bahasa Indonesia?

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

1. *Monolingual resources* yang digunakan pada penelitian ini adalah Tesaurus Bahasa Indonesia tahun 2010.
2. Dataset yang digunakan pada penelitian ini diambil secara manual dari Tesaurus.
3. Fokus utama penelitian ini adalah pembangunan synset
4. Penulis membuat dataset synset yang telah dihasilkan dari tiap kata secara manual. Hal ini dikarenakan susahnya 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. Mengimplementasikan metode komutatif untuk pembangunan synsets Bahasa Indonesia.
2. Mengetahui performansi metode komutatif untuk pembangunan synsets Bahasa Indonesia.

Organisasi Tulisan

Penulisan laporan penelitian tugas akhir ini terdiri dari Studi Terkait, Sistem yang dibangun, Evaluasi, dan Kesimpulan. Pada bagian Studi Terkait berisi teori-teori dan literatur terkait untuk mendukung pengerjaan penelitian tugas akhir. Pada Sistem yang dibangun berisi proses rancangan dan sistem atau produk yang dihasilkan. Selanjutnya Evaluasi berisi hasil pengujian dan analisis hasil pengujian. Kemudian kesimpulan yang didapat dari hasil penelitian berdasarkan uraian pada bagian evaluasi akan dituliskan pada bagian Kesimpulan.

2. Studi Terkait

2.1 WordNet

WordNet merupakan sebuah database kamus Bahasa Inggris yang dikembangkan oleh Princeton Universsity. Perbedaan WordNet dengan kamus bahasa pada umumnya adalah kamus bahasa memfokuskan pada kata sedangkan WordNet memfokuskan diri pada makna kata[1]. 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[1]. Untuk WordNet dalam Bahasa Indonesia, salah satunya telah dikembangkan oleh Universitas Indonesia[8]. 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 Synset

Synset atau synonim set merupakan satuan utama yang digunakan dalam WordNet. Synset adalah kumpulan dari satu atau lebih kata yang memiliki makna yang sama atau sinonim[9]. 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[1]. 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[10]. Hubungan setiap kata dalam synset bersifat komutatif atau simetris, dimana terdapat kata *w1* sinonim dengan kata *w2*, maka *w2* pasti sinonim dengan *w2* pada makna yang sama[1]. 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'[7]. Menurut[1], 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[1][2], selain itu informasi yang terdapat pada Tesaurus telah diakui leksikografer dan merupakan sumber daya yang mudah di unduh dan disediakan oleh Pusat Bahasa Indonesia.

2.4 Gold Standard

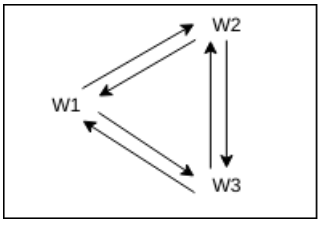
Gold Standard merupakan suatu nilai yang digunakan untuk menganalisis korelasi dari suatu sistem. Standar ini dibuat berdasarkan beberapa pihak ahlli pada bidangnya. Pada Tugas akhir ini, dataset yang digunakan sebagai perbandingan dibuat secara manual secara hati-hati sesuai dengan konsep atau hukum komutatif pada synsets dan juga dicocokkan berdasarkan informasi data yang terdapat pada Kamus Besar Bahasa Indonesia. 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 merupakan 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 *w*1, *w*2, dan *w*3. Tanda panah *w*1 menuju *w*2 menunjukan bahwa *w*1 memiliki sinonim *w*2, begitu juga dengan yang lainnya. Gambar 1 menunjukan sebuah synset harus memiliki keterhubungan komutatif antar semua kata yang ada.

2.6 *F1-Score*

Metode yang digunakan dalam mengukur akurasi sistem pada penelitian tugas akhir ini dengan menggunakan *F1-Score*, untuk mengukur perfomasi menggunakan *F1-Score* dibutuhkan *recall* dan *precision*. Recall yaitu mengukur rasio dari jumlah prediksi yang benar terhadap total prediksi yang diharapkan[11]. Formula untuk *recall* yang umum digunakan pada persamaan 1.

$$r = \frac{tp}{tp + fn} \tag{1}$$

dengan,
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 postif.

Precision yaitu mengukur rasio dari jumlah prediksi yang benar terhadap total prediksi[11]. Formula untuk *precision* yang umum digunakan pada persamaan 2.

$$p = \frac{tp}{tp + fp} \tag{2}$$

dengan,
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 akurasi dalam kasus penelitian ini maka formula disederhanakan. Sehingga formula untuk *precision* pada persamaan 3 dan *recall* pada persamaan 4.

$$precision = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{retrieved\ documents\}|} \tag{3}$$

dan,

$$recall = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{relevant\ documents\}|} \tag{4}$$

dengan,
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 persa-

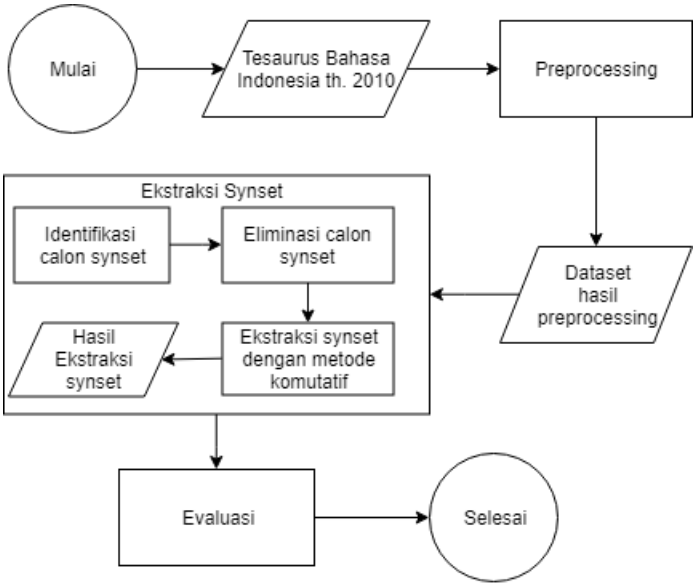
maan 5.

$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{5}$$

3. Sistem yang Dibangun

3.1 Rancangan Sistem

Pada penelitian tugas akhir ini dibangun sebuah sistem yang dapat menghasilkan synset. Metode yang telah diterangkan sebelumnya akan menjadi acuan dalam menentukan pembangunan synset yang sesuai. Rancangan sistem pada penelitian tugas akhir ini dapat dilihat pada gambar (2).



Gambar 2. Flowchart gambaran umum sistem

Gambar (2) menunjukan alur proses yang dilkakukan 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 akurasi sistem dengan menggunakan *F1-Score*.

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[1] dan dapat diunduh secara bebas dan disediakan oleh Pusat Bahasa Indonesia. Dataset yang diunduh merupakan Tesaurus Bahasa Indonesia tahun 2010 dalam format pdf (*Portable Document Format*). Karena dataset berupa pdf tentunya perlu dirubah kedalam format yang dapat dibaca oleh sistem yaitu text, namun untuk merubah formatnya penulis belum menemukan *tool* yang mampu mendapatkan semua item kata secara utuh. Oleh karena itu, karena fokus utama dalam penelitian ini adalah pembangunan synset dengan menggunakan metode komutatif maka item kata yang diambil sebagai data test sebanyak 30 item kata secara manual. Pemilihan item kata Tesaurus 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. suar	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

Data test yang dipilih terdiri dari 30 item kata yang memiliki satu himpunan atau lebih, memiliki sense lebih dari satu, dan terdapat yang tidak memiliki himpunan pada item kata dalam Tesaurus. Selanjutnya 30 item kata yang diambil dari Tesaurus secara manual yang akan dicari synsetnya dengan menggunakan metode komutatif.

3.3 *Preprocessing*

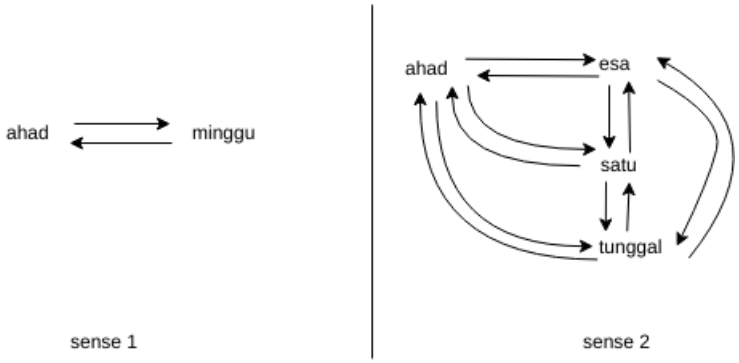
Pada proses pertama pembangunan sistem, dataset Tesaurus melalui proses *preprocessing* yang bertujuan untuk memilih dataset yang akan digunakan dalam proses ekstraksi. Pada proses ini dilakukan pemilahan kata-kata, menghilangkan atribut-atribut yang tidak digunakan seperti *n, v, adv, v, cak, (cak),* dan yang lainnya, karena informasi yang akan diambil hanya himpunan kata yang terdapat pada setiap item kata pada Tesaurus Bahasa Indonesia. 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 synset 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 di atas 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 (3).



Gambar 3. Hubungan komutatif kata 'ahad'.

Pada gambar (3), 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 pembanguan synset.

Algoritma Ekstraksi Synset dengan Metode Komutatif
Setelah proses *preprocessing* dilakukan, selanjutnya adalah proses ekstraksi dengan menggunakan konsep komutatif pada synset. Proses ekstraksi dilakukan dalam beberapa tahap algoritma sebagai berikut. Contohnya adalah kata 'ahad', maka calon synset yang mungkin adalah

1. Identifikasi calon synset yang akan dicari.
   pada tahap ini proses yang dilakukan adalah mencari calon-calon synset yang dapat dihasilkan dari setiap item kata dalam dataset.
   - sense ke satu
     - ahad, minggu
   - sense ke dua
     - ahad, esa
     - ahad, satu
     - ahad, tunggal
     - ahad, esa, satu
     - ahad, esa, tunggal
     - ahad, satu, tunggal
     - ahad, esa, satu, tunggal
2. Tentukan apakah setiap kata pada calon synset memiliki hubungan komutatif.
   Pada tahap ini proses yang dilakukan adalah menentukan calon synset yang memiliki hubungan komutatif. Hubungan komutatif berlaku untuk calon synset yang memiliki lebih dari satu himpunan. Contohnya, akan diambil calon synset 'ahad, satu, tunggal', maka kata 'ahad' dan 'satu' saling komutatif, kata 'ahad' dan tunggal saling komutatif, kata 'satu' dan 'tunggal' saling komutatif agar calon synset tersebut dapat dikatakan valid sebagai synset. Bagi calon synset yang memiliki hubungan komutatif akan ditampung kembali dan bagi yang tidak memiliki hubungan komutatif maka bukan berarti synset yang valid.
3. Eliminasi calon synset yang merupakan subset dari synset yang lainnya.
   Pada tahap ini proses yang dilakukan adalah eliminasi calon synset yang merupakan subset dari synset yang lainnya. Contohnya adalah pada sense ke dua terdapat tujuh calon synset dan semua calon synset merupakan synset yang valid dengan konsep komutatif. Maka setiap calon synset yang merupakan subset dari synset yang lainnya akan dihapus atau dieliminasi.
4. Ambil sisa dari eliminasi calon synset.
   Pada proses ini akan diambil calon synset yang tersisa dan dijadikan sebagai synset. Contohnya pada kata 'ahad' synset akhir yang dihasilkan adalah
   - sense ke satu
     - ahad, minggu
   - sense ke dua
     - ahad, esa, satu, tunggal

3.5 Perhitungan Evaluasi

Proses terakhir yang dilakukan adalah perhitungan akurasi yang berguna untuk mengukur perfomasi dari sistem yang telah dibuat. Proses evaluasi menggunakan formula *F1-Score* yang telah dijelaskan sebelumnya. *Recall* diambil dari nilai data hasil prediksi synset, sedangkan *precision* berasal dari data prediksi sebenarnya yang didapatkan dari hasil komutatif secara manual. Proses yang dilakukan dalam pembuatan *gold standard* menggunakan metode yang sama yaitu konsep komutatif pada synset.

4. Evaluasi

4.1 Skenario Pengujian

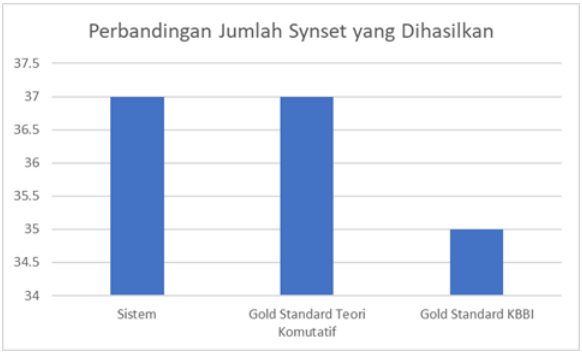
Dalam menguji metode yang digunakan dalam penelitian tugas akhir ini, dilakukan skenario pengujian sebagai berikut.

1. Skenario - 1 (Evaluasi implementasi metode komutatif dengan teori komutatif).
   Pada skenario pengujian ini hasil implementasi metode komutatif yang menhasilkan synset dari program akan dibandingkan dengan synset yang dihasilkan secara manual dengan menggunakan teori komutatif.

2. Skenario - 2 (Evaluasi implementasi metode komutatif dengan synset perbandingan KBBI).
   Pada skenario pengujian ini hasil synset program dibandingkan dengan synset yang dihasilkan secara manual dengan teori komutatif dan dicocokan dengan informasi yang terdapat pada Kamus Besar Bahasa Indonesia (KBBI) online. Selain itu, terdapat hasil *upper bound* dan *lower bound*. *Upper bound* digunakan sebagai nilai akurasi tertinggi, dimana proses mendapatkan *upper bound* dengan membandingkan hasil synset yang dibuat secara manual. Pada penelitian ini synset yang dibuat secara manual dibagi menjadi dua, yang pertama hasil synset dengan menggunakan konsep komutatif dan yang kedua adalah hasil synset dengan konsep komutatif dan digabungkan dengan informasi kata pada Kamus Besar Bahasa Indonesia. *Lower bound* digunakan sebagai nilai akurasi terendah. Untuk mendapatkan nilai *lower bound* dibagi menjadi dua data, yang pertama synset yang diambil dari sinonim yang terdapat pada kata dalam Tesaurus dan dibandingkan dengan hasil synset dengan konsep komutatif dan digabungkan dengan informasi kata pada Kamus Besar Bahasa Indonesia.

4.2 Analisis dan Hasil Pengujian

Setelah dilakukan pengujian sistem menggunakan dua skenario pengujian, berikut disajikan hasil pengujian pada setiap skenario beserta analisisnya.



Gambar 4. Hasil Jumlah Synset yang Dihasilkan.

Dari Gambar 4, synset yang dihasilkan dari 30 data pengujian sebanyak 37 synset yang didapatkan dari sistem, sedangkan untuk gold standard yang dihasilkan dengan mencocokkan dengan KBBI dihasilkan synset sebanyak 35 synset. Dari 30 dataset mampu menghasilkan lebih dari satu synset.

Tabel 3. Tabel hasil evaluasi implementasi metode komutatif dengan teori komutatif.

	Precision %	**Recall %**	**F1 %**
This work*	100%	100%	100%

Tabel 4. Tabel hasil evaluasi implementasi metode komutatif dengan synset perbandingan KBBI.

System	**Precision %**	**Recall %**	**F1 %**
This work*	**67.56**	**71.42**	**69.44**
Upper bound	67.56	71.42	69.44
Lower bound	43.58	48.57	45.94

Sistem yang dibangun pada penelititan ini menghasilkan performansi dari implementasi metode komutatif dengan perbandingan metode komutatif menghasilkan nilai sebesar 100%. Sedangkan untuk perbandingan evaluasi dengan synset hasil yang dicocokkan dengan KBBI menghasilkan performansi sebesar 69.44%. Hasil yang didapatkan dari sistem yang dibandingkan dengan hasil synset teori komutatif tinggi karena sistem mampu menghasilkan synset sesuai dengan implementasi metode komutatif.

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. Setiap kata atau dataset dapat memiliki lebih dari satu synset.
2. Performansi yang dihasilkan dari implementasi metode komutatif dengan gold standard teori komutatif menghasilkan nilai 100%
3. 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 pengujian untuk melakukan pengujian.
2. Menggunakan dataset Tesaurus Bahasa Indonesia versi terbaru.

Daftar Pustaka

[1] Gunawan. *Akuisisi Gloss Berbasis Ekstraksi Synonim Set Menggunakan Supervised Learning*. Institut Teknologi Sepuluh November, 2016.

[2] Andy 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.

[3] Christiane Fellbaum. *WordNet*. Wiley Online Library, 1998.

[4] Mortaza Montazery and Heshaam Faili. Automatic persian wordnet construction. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 846–850. Association for Computational Linguistics, 2010.

[5] Changki Lee, Geunbae Lee, and Seo Jung 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.

[6] Tim Redaksi Kamus Bahasa Indonesia. Kamus bahasa indonesia. *Jakarta: Pusat Bahasa Departemen Pendidikan Nasional*, 2008.

[7] Eko Endarmoko. *Tesaurus Bahasa Indonesia*. Gramedia Pustaka Utama, 2007.

[8] Desmond Darma Putra, Abdul Arfan, and Ruli Manurung. Building an indonesian wordnet. In *Proceedings of the 2nd International MALINDO Workshop*, pages 12–13, 2008.

[9] Catherine Havasi, Robert Speer, and Jason Alonso. Conceptnet 3: a flexible, multilingual semantic network for common sense knowledge. In *Recent advances in natural language processing*, pages 27–29. Citeseer, 2007.

[10] George A Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J Miller. Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, 3(4):235–244, 1990.

[11] Jérôme Euzenat. Semantic precision and recall for ontology alignment evaluation. In *IJCAI*, volume 7, page 348353, 2007.

Lampiran

Berikut merupakan lampiran-lampiran pada penelitian tugas akhir ini.

Tabel 5. 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’]]]