

Pembangunan *Synonym Set* untuk WordNet Bahasa Indonesia dengan Menggunakan Metode Komutatif

Dina Juni Restina ^{#1}, Moch. Arif Bijaksana ^{*2}, Kemas Muslim Lhaksana ^{#3}

[#] *Department of Informatics, Telkom University
Bandung, West Java, Indonesia*

¹ dinajr@student.telkomuniversity.ac.id

² arifbijaksana@telkomuniversity.ac.id

³ kemasmuslim@telkomuniversity.ac.id

Abstract

In NLP (Natural Language Processing) many lexical semantic linkages can be applied, for example is the WordNet application. The WordNet has been built using a commutative method in Developing synset (synonym set). The synonym set has been built has a commutative attribute which means that if a word1 has a synonym of word2, then word2 must also have a synonym of word1, it can be said that the synonym set is valid, such as the implementation of Princeton WordNet. WordNet was first created in the English language, before the researchers translated into various languages such as Japanese, Arabic, Turkish and Indonesian and other languages. For this reason, it takes the development of WordNet to participate in helping other researchers, so in the future The Indonesian WordNet would have more complete vocabulary. WordNet has been built focusing on synset extraction, which is the first stage of WordNet development before the stages of relations between synset and gloss words. The development of synonym set uses the Indonesian Thesaurus dictionary as the source of words. The value of F-measure from the development of synonym set using the commutative method is 66 percent which is the value of the evaluation of the suitability of the synset produced by the system and the synset made by linguists based on the specified gold standard.

Keywords: Commutative method, Synset dan WordNet Bahasa Indonesia.

Abstrak

Dalam NLP (*Natural Language Processing*) banyak keterkaitan semantik leksikal yang dapat diaplikasikan, salah satunya adalah aplikasi WordNet. WordNet yang dibangun menggunakan metode komutatif dalam proses pembangunan *synonym set (synset)*-nya. *Synset* yang akan dibangun harus memiliki sifat komutatif baru dapat dikatakan *synset* yang bernilai valid, yang berarti jika sebuah kata w_1 memiliki sinonim w_2 , maka w_2 juga harus memiliki sinonim w_1 , seperti yang terjadi pada *Princeton WordNet*. WordNet pertama kali dibuat dalam bahasa Inggris, sebelum para peneliti menerjemahkan kedalam berbagai bahasa seperti bahasa Jepang, bahasa Arab, bahasa Turki serta bahasa Indonesia dan bahasa lainnya. Untuk itu dibutuhkan pembangunan WordNet untuk turut serta membantu peneliti lain agar kedepannya WordNet Bahasa Indonesia yang ada memiliki kosa kata yang lebih lengkap. WordNet yang akan dibangun akan berfokus pada ekstraksi synset yaitu tahapan pertama pembangunan WordNet sebelum tahapan relasi antar *synset* dan gloss kata. Pembangunan *synset* ini menggunakan Tesaurus Bahasa Indonesia sebagai sumber kata. Nilai *F-measure* dari pembangunan *synset* dengan menggunakan metode komutatif adalah sebesar 66 persen yang merupakan nilai evaluasi kesesuaian *synset* yang dihasilkan oleh sistem dan *synset* yang dibuat oleh ahli bahasa berdasarkan *Gold Standard* yang ditentukan.

Kata Kunci: Metode komutatif, *Synset* dan WordNet Bahasa Indonesia.

I. PENDAHULUAN

WORDNET adalah sebuah database kamus bahasa Inggris yang pertama kali dikembangkan oleh Princeton University. Sedangkan menurut Miller, WordNet merupakan pangkalan data leksikal elektronik. Dalam WordNet, kata benda, kata kerja, kata sifat, dan kata keterangan dikelompokkan dalam suatu set sinonim yang dinamakan *synonym set* (*synset*). Suatu *synset* saling terkait dengan *synset* lainnya [1]. Umumnya kamus bahasa adalah kamus yang memiliki fokus pada makna kata sedangkan WordNet yang sedang dibangun memfokuskan diri kepada persamaan makna kata (sinonim). Kumpulan kata dalam WordNet dapat merepresentasikan satu makna atau bisa disebut dengan *synset*, selain dari representasi makna, sebuah kata juga terdapat hubungan antar kata seperti hipernim, hiponim, holonim, meronim, dan lain-lain. Bagi komputer WordNet berfungsi sebagai sumber untuk melakukan analisis semantik pada teks bahasa alami yang kemudian akan memiliki sebuah output berupa *synset* (*synonym set*).

Pembangunan WordNet pertama kali dilakukan oleh Princeton University. Keberhasilan Princeton University dalam membangun WordNet bahasa Inggris kemudian menjadi awal bagi para peneliti dari berbagai negara untuk turut mengembangkan WordNet yang ditranslasi kedalam banyak bahasa di dunia. terbukti hingga saat ini telah ada setidaknya 40 atau lebih WordNet dalam bahasa asing baik yang telah selesai ataupun sedang dikerjakan [2]. Negara - negara yang turut menerjemahkan WordNet kedalam bahasa mereka diantaranya yaitu Jepang [3], Arab [4] Turki [5]. WordNet juga diterjemahkan kedalam bahasa yang telah mati seperti bahasa Sansekerta dan bahasa Latin [6]. Dengan banyaknya WordNet yang ada saat ini menyebabkan banyak penelitian didalam bidang *Natural Language Processing* (NLP) merasa terbantu dimana penelitian tersebut umumnya membutuhkan kamus, daftar kata ataupun makna kata contohnya adalah penggunaan WordNet pada penelitian *information retrieval* [7] [8], *question and answering system* [9], hingga pengaplikasian WordNet pada penelitian klasifikasi teks [10]. Hal itu membuktikan bahwa dengan adanya WordNet dapat membantu menunjang basis pengetahuan dibidang kecerdasan buatan. Jika WordNet merupakan jaringan semantik yang sangat besar maka *synset* digambarkan sebagai sebuah node. WordNet yang telah dikerjakan sebelumnya telah banyak membahas mengenai pebangunan WordNet itu sendiri, akan tetapi masih sedikit yang berfokus pada pembangunan *synset*, padahal *synset* memegang peranan penting dalam menentukan baik atau buruknya *output* suatu WordNet. Untuk itu Penelitian yang dibangun berfokus pada ekstraksi *synset* yaitu tahapan pertama pembangunan WordNet sebelum tahapan relasi antar *synset* dan gloss kata. Proses ekstraksi *synset* untuk menghasilkan *synset* yang bernilai valid akan menggunakan metode Komutatif seperti pada PWN (*Princeton WordNet*), artinya jika sebuah kata w_1 memiliki sinonim w_2 maka w_2 juga harus merupakan sinonim w_1 . Dikarenakan pada kenyataanya relasi komutatif seperti ini tidak selalu terjadi dalam Tesaurus Bahasa Indonesia [11]. Pembangunan WordNet bahasa Indonesia ini menggunakan Tesaurus sebagai sumber kata.

II. PENELITIAN TERKAIT

Jurnal yang dipublikasikan oleh Hendrik dan Andhik Budi Cahyono membahas mengenai Pembangunan WordNet Bahasa Indonesia berbasis Linked Data, adapun tahapan pengembangan meliputi identifikasi sumber data, ekstraksi data, transformasi data, pemuatan data ke dalam basis data relasional, serta pemetaan basis data relasional ke model RDF [12]. Proses pemetaan menggunakan *framework* D2RQ dan menghasilkan WordNet Bahasa Indonesia berbasis Linked Data. Makalah ini telah berhasil menyediakan data set WordNet Bahasa Indonesia dalam format *machine understandable*. Menggunakan format ini, pengguna dapat melakukan otomatisasi proses daripada dengan menggunakan WordNet model tradisional. Meskipun demikian, dalam makalah ini belum dilakukan pengujian pemanfaatan data set untuk berbagai kasus aplikasi penelitian di bidang NLP.

Jurnal dengan judul *Building an Indonesian WordNet* yang dipublikasikan oleh Desmond Darma Putra, dkk., membahas tentang pembangunan WordNet Bahasa Indonesia menggunakan pendekatan dengan *expand model* yaitu dengan memetakan entri WordNet Indonesia berdasarkan definisi kata dari PWN [2]. Pada jurnal ini dibahas satu *synset* PWN dapat bersesuaian dengan banyak definisi pada KBBI. Akan tetapi secara teori ini dapat terjadi sebaliknya seperti definisi arti kata KBBI dapat dipetakan ke beberapa *synset*

di PWN akan tetapi seharusnya hal itu tidak banyak terjadi melihat *synset* sudah mewakili semantik yang berbeda entitas. Namun demikian, perbedaan semantik seperti itu tidak sepenuhnya universal, dan hal itu memang telah ditunjukkan di beberapa tingkat polisemi WordNet yang terlalu halus atau dengan kata lain perbedaan makna kata yang sangat dekat. Kelemahan dari pendekatan ini adalah dapat mengacaukan *synset* pada saat menggabungkan beberapa *synset* menjadi satu *synset*.

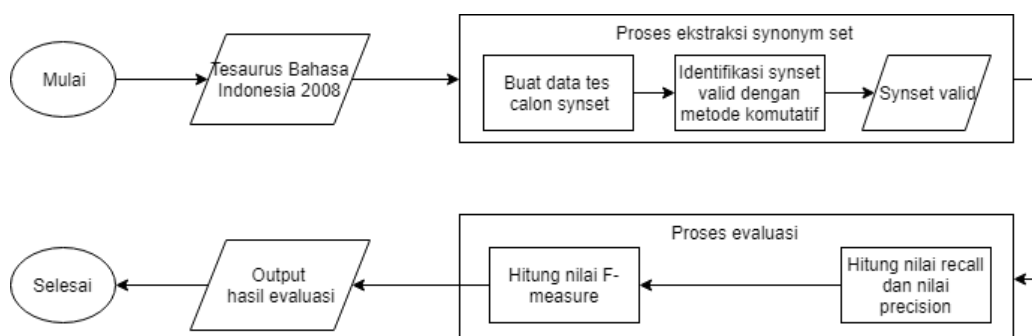
Gunawan dan Andy Saputra menjelaskan melalui jurnalnya mengenai penelitian pembangunan *synset* dengan menggunakan pendekatan *Monolingual Lexical Resources* [13]. Dan pendekatan utamanya menggunakan Tesaurus dan KBBI sebagai sumber utama dari data tesnya. Pada dasarnya, jurnal ini memaparkan langkah - langkah dalam pembangunan *synset* menggunakan pendekatan *Monolingual Lexical Resources* yaitu mengekstraksi kandidat sinkronisasi dari *entri Tesaurus*, menambahkan lebih banyak kandidat *synset* dari KBBI, menghilangkan *redundant* atau kata yang mirip dan terakhir, menggunakan teknik pengelompokan, menggabungkan kandidat *synsets* yang dianggap memiliki kesamaan arti. Akan tetapi jurnal ini memiliki kelemahan yaitu pada proses pengelompokan (*merging*), teknik pengelompokan pada jurnal ini masih berdasarkan kategori sehingga pengelompokan yang dilakukan kurang tepat untuk beberapa kasus seperti jika terdapat data dari KBBI dan Tesaurus tidak relevan untuk satu sama lain dan ada beberapa data yang ada di KBBI dan Tesaurus, yang merupakan kata bermakna tunggal, tetapi memiliki kelas semantik yang berbeda, meskipun jika kita lihat secara manual, mereka merujuk ke lemma yang sama dan mereka seharusnya memiliki kelas semantik yang sama. Serta penelitian ini tidak melibatkan ahli dalam Bahasa, sehingga hasil penelitian ini dianggap sebagai versi prototipe dari pembangunan *synset* untuk WordNet Bahasa Indonesia.

III. PERANCANGAN SISTEM

A. Rancangan Sistem

Rancangan sistem yang dibuat dalam penelitian ini dalam membangun *synset* untuk WordNet Bahasa Indonesia yang akan menghasilkan *synset* valid yang bersifat komutatif adalah sebagai berikut.

1) *Flowchart sistem*: Adapun alur proses dari gambaran sistem secara umum seperti terlihat pada Gambar 1



Gambar 1. Flow Chart System

Gambar 1 menunjukkan alur dari proses jalannya sistem yaitu dimulai dari proses pemilihan kata untuk data tes calon *synset* dari Tesaurus Bahasa Indonesia, selanjutnya dilakukan proses ekstraksi *synset* yaitu proses identifikasi data tes calon *synset* dengan menggunakan metode komutatif untuk menghasilkan *synset* yang valid. Setelah proses ekstraksi *synset* maka sistem akan menghitung nilai *recall* dan nilai *precision* untuk menghasilkan nilai *F-measure* sebagai *output* nilai evaluasi sistem.

B. WordNet

WordNet adalah hasil proyek penelitian di Princeton University yang bertujuan untuk memodelkan pengetahuan leksikal pembicara asli Bahasa Inggris. Informasi di dalam WordNet diorganisasikan ke

dalam kelompok logikal yang disebut *synset*. Tiap-tiap *synset* berisikan bentuk synonym kata dan pointer semantik yang menjelaskan hubungan antara satu *synset* dengan *synset* lainnya. WordNet adalah basis data leksikal *online* yang menyediakan tempat penyimpanan leksikal bahasa. WordNet didesain untuk menyediakan hubungan antara empat label kelas bagian perkataan (*Parts of Speech*, POS) *-noun, verb, adjective* dan *adverb*. Unit terkecil WordNet adalah *synset* yang merepresentasikan makna spesifik sebuah kata yang nantinya akan berelasi dari satu lema dengan lema lainnya [1].

C. Tesaurus

Kata tesaurus berasal dari kata thesauros, bahasa Yunani, yang bermakna 'khazanah'.Lambat laun, kata tersebut mengalami perkembangan makna, yakni 'buku yang dijadikan sumber informasi'. Tesaurus berisi seperangkat kata yang saling bertalian maknanya. 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. Misalnya, pencarian kata lain untuk kata hewan, pengguna tesaurus dapat mencarinya pada lema '**hewan**' [14]. Seperti terlihat pada Gambar 2

hewan n binatang, dabat, fauna, sato, satwa

Gambar 2. Pencarian lema hewan pada tesaurus

D. Metode Komutatif

Synset dikatakan valid jika memiliki sifat komutatif. Seperti pada WordNet Princeton University, relasi sinonim seharusnya komutatif yang artinya jika sebuah kata w_1 memiliki sinonim w_2 , maka w_1 harus merupakan sinonim dari w_2 [11].

Pencarian *synset* yang bernilai valid dilakukan dengan menggunakan tabel matriks. Seperti terlihat pada Tabel I untuk tabel matriks kata *fana*. *Sense* kata *fana* yaitu *sementara, kontemporer*, dan kata *fana* itu sendiri, berfungsi sebagai baris dan kolom.

Tabel I
TABEL MATRIKS KATA FANA

	<i>fana</i>	<i>sementara</i>	<i>temporer</i>
<i>fana</i>	T	T	T
<i>sementara</i>	T	T	T
<i>temporer</i>	F	T	T

Penjelasan mengenai Tabel I ialah *cell* yang bernilai *true* menunjukkan bahwa kata baris dan kolom yang diwakilkanya bernilai komutatif dan untuk *cell* yang bernilai *false* maka sebaliknya. Contohnya pada baris *temporer* dan kolom *fana* menunjukkan *cell* bernilai *false*, hal ini berarti kata *fana* bukan termasuk *sense* dari kata *temporer*. Dari tabel matriks diatas maka dihasilkan satu buah *synset* yaitu [*fana,sementara*]. Namun ada kata jika diterapkan pada matriks kata dapat menghasilkan lebih dari satu *synset*, hal ini disebabkan adanya nilai *false* pada tabel matriks. Jika pada tabel matriks kata *fana* kata *temporer* otomatis hilang, itu dikarenakan kata *fana* yang merupakan kata kunci dari tabel bukan merupakan *sense* dari *temporer*. Akan tetapi jika nilai *false* pada tabel bukan untuk kata kunci, itu menyebabkan *cell* dengan dua kata yang diwakilkanya tidak dapat berada dalam satu *synset*.

E. Synonym Set

Synset adalah sebuah himpunan yang tersusun dari satu atau lebih kata yang memiliki hubungan sinonim, di mana masing-masing anggotanya dapat saling menggantikan dalam sebagian besar penggunaan kata dalam kalimat tanpa mengubah makna [11]. Seperti *synset* yang dihasilkan dari matriks kata *fana*. Lihat pada Tabel II

Tabel II
PENERAPAN SYNSET PADA KALIMAT

Synset	Kalimat
[<i>fana,sementara</i>]	Semua yang ada dibumi ini adalah <i>fana</i>
	Semua yang ada dibumi ini adalah <i>sementara</i>

F. Gold Standard

Gold Standard merupakan nilai yang digunakan untuk mengevaluasi korelasi dari suatu sistem. Standar ini dibuat oleh manusia yang merujuk pada para ahli dibidang leksikal. Gold Standard diterapkan untuk evaluasi korelasi sistem ialah dengan menerapkan himpunan kata pada *synset* yang didapat dari Tesaurus Bahasa Indonesia dalam satu kalimat dan himpunan kata pada *synset* tersebut juga dicari maknanya menggunakan Kamus Besar Bahasa Indonesia (KBBI) [15] [16]. *Synset* yang telah sesuai ialah jika tiap himpunan kata pada satu *synset* dapat saling menggantikan penggunaannya pada sebuah kalimat, tanpa menghilangkan makna kalimat tersebut.

G. Data Uji

Data yang digunakan sebagai data uji dalam penelitian merupakan kata yang diambil dari Tesaurus Bahasa Indonesia dalam bentuk pdf yang diterbitkan tahun 2008 [14]. Data yang diambil ialah sebanyak 50 kata yang kemudian akan diekstrak untuk menghasilkan *synset* yang valid. Data Uji seperti terlihat pada Tabel III

Tabel III
DAFTAR KATA DARI TESAUROS BAHASA INDONESIA

Daftar Kata dari Tesaurus Bahasa Indonesia				
<i>acuh</i>	<i>lem</i>	<i>atlit</i>	<i>lamaran</i>	<i>pundak</i>
<i>eka</i>	<i>gaji</i>	<i>oral</i>	<i>jahilliah</i>	<i>bajul</i>
<i>eks</i>	<i>handuk</i>	<i>masehi</i>	<i>edan</i>	<i>bantahan</i>
<i>eksponen</i>	<i>idas</i>	<i>nagari</i>	<i>otot</i>	<i>cabuk</i>
<i>aborsi</i>	<i>idealis</i>	<i>mambang</i>	<i>rambut</i>	<i>cadar</i>
<i>fana</i>	<i>pahala</i>	<i>maharani</i>	<i>kafilah</i>	<i>dam</i>
<i>faksi</i>	<i>padas</i>	<i>dandang</i>	<i>fiber</i>	<i>pacul</i>
<i>kabin</i>	<i>harap</i>	<i>darat</i>	<i>fiksi</i>	<i>orasi</i>
<i>jambe</i>	<i>warkat</i>	<i>abrasi</i>	<i>cahar</i>	<i>galat</i>
<i>ialah</i>	<i>madukara</i>	<i>gagu</i>	<i>kabung</i>	<i>mampir</i>

Kata yang telah dipilih kemudian akan diproses oleh sistem dan akan menghasilkan satu atau lebih *synset*. Daftar kata dari Tesaurus Bahasa Indonesia dipilih secara *random*.

H. Proses Ekstraksi *synset*

Proses ekstraksi dalam pembangunan *synset* untuk WordNet Bahasa Indonesia dalam penelitian ini menggunakan metode komutatif. Adapun tahapan algoritma ekstraksi *synset* dengan metode komutatif adalah sebagai berikut :

- 1) Daftar kata yang telah dipilih kemudian akan dicari *sense*-nya dengan menggunakan Tesaurus Bahasa Indonesia. Lalu tiap *sense* tersebut akan dicari kembali himpunan sinonimnya. Setelahnya seluruh himpunan-himpunan sinonim tersebut akan disimpan kedalam "data_synonim.txt", baru kemudian diproses untuk mencari *synset* yang bersifat valid.
- 2) Setelah data tes dibuat, barulah proses ekstraksi *synset* berjalan. Dengan langkah pertama yaitu untuk kata *fana* memiliki *sense* seperti yang terlihat pada Gambar 3

sementara	temporer
-----------	----------

Gambar 3. Sense kata fana

Kedua *Sense* dari kata *fana* kemudian akan dicari himpunan sinonimnya. Seperti yang terlihat pada Tabel IV

Tabel IV
SINONIM DARI *Sense* KATA FANA

Kata	Sinonim
<i>fana</i>	<i>sementara,temporer</i>
<i>sementara</i>	<i>selagi,selama,senyampang,tatkala,tengah,darurat,fana,temporer</i>
<i>temporer</i>	<i>darurat,provisional</i>

- 3) Langkah selanjutnya adalah mencari sinonim yang memiliki *Keyword* : *fana*, yaitu *sementara*, untuk kemudian kata ini dan *fana* itu sendiri akan dijadikan kolom dan baris pada matriks.
- 4) Langkah ketiga ialah mencari potensial *sense*. potensial *sense* adalah *sense* yang menjadi kandidat dari *synset* valid. Adapun langkah mencari potensial *sense* dapat dilihat pada Algorithm 1

Result: Potensial sense

rows = sense kata;

cols = sense kata;

potential_sense = {};

for row in rows **do**

for col in cols **do**

if row != col **then**

 periksa apakah col merupakan sense dari row;

 jika iya tambahkan kedalam list potensial_sense;

else

 tampilkan col;

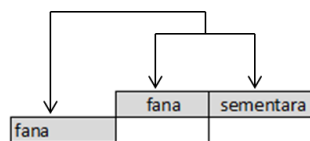
end

end

end

Algorithm 1: Mencari Potensial Sense

Tahapan untuk memastikan bahwa *synset* yang dihasilkan bernilai valid maka dilakukan pengecekan dengan membandingkan baris dan kolom pada tabel matriks. Pengecekan ini dilakukan untuk tiap baris untuk memastikan apakah kolom merupakan *sense* dari baris. Lihat pada Gambar 4



Gambar 4. Mencari Potensial *Sense*

Potensial *sense* yang dihasilkan dari langkah 4 adalah sebagai berikut :

```
{
  'fana' : ['sementara'],
  'sementara' : ['fana']
}
```

- 5) Langkah terakhir ialah validasi *synset*. Langkah validasi *synset* ini bertujuan untuk memastikan *synset* yang dihasilkan oleh sistem bersifat komutatif. Adapun algoritmanya dapat dilihat pada Algorithm 2

Result: Synset yang valid
 synset_to_list = list dari synset;
 sinonimity_set = [];
for keyword1 in synset_to_list **do**
 | l1 <- synset_to_list[keyword1];
 | **for** word1 in l1 **do**
 | | synonyms = kata sinonim word1;
 | | **for** word2 in l1 **do**
 | | | **if** word1 == word2 **then**
 | | | | Periksa tiap kata1 dalam list1 merupakan sense kata2;
 | | | | dan kata2 merupakan sense kata1;
 | | | **end**
 | | | **if** word2 not in synonyms **then**
 | | | | hapus word2 dari l1;
 | | | **end**
 | | **end**
 | **end**
 | **if** l1 not in sinonimity_set **then**
 | | sinonimity_set =+ 1
 | **end**
end
if sinonimity_set tidak ada **then**
 | tampilkan keyword
end

Algorithm 2: Validasi Synonym Set

Synset bernilai valid yang dihasilkan pada langkah 5 adalah sebagai berikut :

['fana' , 'sementara']

1. Hasil Evaluasi

Hasil evaluasi akan dihasilkan melalui proses membandingkan *synset* yang dihasilkan oleh sistem dengan penerapan materi komutatif dan *synset* dari ahli bahasa maka akan dihitung performansi dari sistem (*F –measure*) dengan menggunakan *recall* dan *precision*.

Perhitungan akurasi sistem membutuhkan nilai *recall* dan nilai *precision*. Menurut Sulisty-Basuki [17], Rasio perolehan (*recall*) adalah perbandingan dokumen ditemukan dengan jumlah total dokumen relevan dalam sistem. Sedangkan rasio ketepatan (*precision*) adalah perbandingan antara dokumen relevan dengan jumlah dokumen yang ditemu balik dalam penelusuran.

Perolehan (*recall*) berhubungan dengan kemampuan sistem untuk memanggil dokumen yang relevan [17]. Untuk menghitung nilai perolehan (*recall*) digunakan persamaan sebagai berikut:

$$Recall (R) = \frac{JumlahDokumenRelevanyangTerambil}{JumlahDokumenRelevandalamDatabase} \quad (1)$$

atau

$$Recall (R) = \frac{TP}{TP + FN} \quad (2)$$

Ketepatan (*Precision*) berkaitan dengan kemampuan sistem untuk tidak memanggil dokumen yang tidak relevan. Untuk menghitung nilai ketepatan (*Precision*) digunakan persamaan sebagai berikut [17].

$$Recall (R) = \frac{JumlahDokumenRelevanyangTerambil}{JumlahDokumenRelevandalamPencarian} \quad (3)$$

atau

$$Precision (P) = \frac{TP}{TP + FP} \quad (4)$$

Persamaan diatas memiliki keterangan yaitu *TP* merupakan hasil temu dokumen yang sesuai dengan yang diharapkan oleh user, *FP* merupakan dokumen yang ditampilkan oleh sistem dengan nilai sebenarnya adalah negatif, dan *FN* merupakan dokumen yang tidak ditampilkan oleh sistem dengan nilai sebenarnya adalah positif [17].

Setelah nilai *Recall* dan nilai *Precision* ditemukan, sistem akan menghitung nilai *F-measure* [18] dengan persamaan sebagai berikut :

$$F\text{-measure} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

IV. HASIL DAN DISKUSI

A. Skenario Pengujian

Pengujian yang dilakukan adalah dengan membandingkan *synset* hasil sistem yang telah diterapkan materi komutatif untuk metode komutatif dengan hasil evaluasi yaitu 100 persen. *Output* dari sistem kemudian akan dibandingkan dengan *synset* dari ahli bahasa yang dibuat berdasarkan *Gold Standar*. Seperti terlihat pada Tabel VI

Tabel V
CONTOH PERBANDINGAN DATA SYNSET MANUAL DAN SYNSET AHLI BAHASA

Kata	Output Sistem	Synset dari Ahli Bahasa
<i>fana</i>	[<i>fana,sementara</i>]	[<i>fana,sementara</i>]
<i>pundak</i>	[<i>bahu,pundak</i>]	[<i>bahu,pundak</i>]
<i>buaya</i>	[<i>bajul,buaya</i>]	[<i>bajul,buaya</i>]
<i>bantahan</i>	[<i>bantahan,sanggahan,sangkalan</i>], [<i>bantahan,sanggahan,tentangan</i>]	[<i>bantahan,sanggahan,sangkalan,tentangan</i>]
<i>cabuk</i>	[<i>cabuk,kelemayuh,pekung</i>]	[<i>cabuk,kelemayuh,pekung</i>]
<i>cadar</i>	[<i>cadar,kerudung,selubung</i>]	[<i>cadar</i>]
<i>dam</i>	[<i>bendungan,dam,tanggul,waduk</i>]	[<i>bendugan,dam,tanggul,waduk</i>]
<i>dandang</i>	[<i>belanga,dandang</i>]	[<i>belanga,dandang</i>]
<i>darat</i>	[<i>bumi,darat,tanah</i>]	[<i>bumi,darat,tanah</i>]
<i>aborsi</i>	[<i>aborsi,engguguran</i>]	[<i>aborsi,engguguran</i>]

B. Hasil Pengujian

Setelah dilakukan skenario pengujian maka didapatkan hasil pengujian yaitu jumlah *synset* yang dihasilkan oleh ahli bahasa adalah 51 buah dan *synset* yang dihasilkan oleh sistem adalah 73 buah sehingga dengan menggunakan persamaan (2) menghasilkan nilai *Recall*, dalam menghitung nilai *Precision* digunakan persamaan (4) dan dalam menghitung nilai *F-measure* digunakan persamaan (5), nilai dari *recall*, *precision* dan *F-measure* dapat dilihat pada Tabel VI

Tabel VI
HASIL PENGUJIAN SYNSET HASIL SISTEM DAN SYNSET AHLI BAHASA

<i>Recall</i>	<i>Precision</i>	<i>F-measure</i>
70 %	62 %	66 %

Berdasarkan hasil pengujian pembangunan *synset* menggunakan metode komutatif dengan data tes yang telah ditentukan. Dapat dilihat pada Tabel VI, didapatkan hasil *recall* sebesar 70 persen, nilai

precision sebesar 62 persen dan nilai *F-measure* sebesar 66 persen. Perolehan nilai dapat dipengaruhi oleh beberapa faktor yaitu dapat dipengaruhi oleh perbedaan jumlah *synset* yang dihasilkan sistem dan jumlah *synset* dari ahli bahasa. Hal ini dapat kita lihat pada Tabel V. Contoh dari kata *bantahan*, *synset* yang dihasilkan oleh sistem berjumlah dua buah *synset* yaitu [*bantahan, sanggahan, sangkalan*] dan [*bantahan, sanggahan, tentangan*] sedangkan *synset* yang dihasilkan oleh ahli bahasa berjumlah satu buah *synset* yaitu [*bantahan, sanggahan, sangkalan, tentangan*]. Sehingga menyebabkan perbedaan jumlah *synset* hasil sistem dan jumlah *synset* dari ahli bahasa, dimana jumlah *synset* yang dihasilkan sistem lebih banyak bila dibandingkan dengan jumlah *synset* dari ahli bahasa.

Perolehan *synset* dari kata *cadar* juga dapat mempengaruhi perolehan nilai pada sistem. Karena untuk *synset* yang dihasilkan sistem dari kata *cadar* adalah [*cadar, kerudung, selubung*] dan *synset* yang diperoleh dari ahli bahasa adalah kata [*cadar*] itu sendiri. Sehingga dapat dikatakan, dari sekian banyak data tes yang dipilih terdapat beberapa kata yang akan menghasilkan *synset* yang berbeda antara sistem dan ahli bahasa.

Beberapa faktor yang telah disebutkan akan berpengaruh pada nilai *recall* dan nilai *precision*. Mengingat *recall* merupakan kemampuan sistem dalam memanggil dokumen yang relevan atau dapat dikatakan bahwa nilai *recall* yang dihasilkan dipengaruhi oleh perbandingan jumlah *synset* yang diprediksi benar oleh sistem dengan jumlah *synset* yang ditebak oleh sistem, dan nilai *precision* adalah perbandingan antara dokumen relevan dengan jumlah dokumen yang ditemu balik dalam penelusuran atau dapat dikatakan nilai *precision* yang dihasilkan adalah nilai yang dipengaruhi oleh perbandingan jumlah *synset* yang dinyatakan benar oleh sistem dengan banyak *synset* yang tersedia pada WordNet. Dan untuk nilai *F-measure* dipengaruhi oleh nilai *recall* dan *precision*.

V. KESIMPULAN

Hasil dari penelitian ini adalah sistem menghasilkan nilai *F-measure* sebesar 66 persen. Nilai *F-measure* yang dihasilkan sistem merupakan nilai evaluasi kesesuaian *synset* yang dihasilkan oleh sistem dan *synset* yang dibuat oleh ahli bahasa dengan menggunakan *Gold Standard* yang ditentukan. Penelitian ini belum bisa dibandingkan dengan penelitian lain dikarenakan sejauh ini peneliti belum menemukan penelitian sejenis untuk WordNet Bahasa Indonesia yang berfokus pada tahapan pertama dalam membangun WordNet bahasa Indonesia yaitu tahapan ekstraksi *synsets* dengan menggunakan metode komutatif.

Penggunaan daftar kata pada penelitian yang akan dilakukan kedepannya diharapkan akan bertambah. Dan sebaiknya pencarian data tes peneliti yang bersumber dari Tesaurus Bahasa Indonesia [14] memiliki edisi yang sama dengan Kamus Besar Bahasa Indonesia (KBBI) [15] [16] yang digunakan oleh ahli bahasa dalam pembangunan *synset*. Dikarenakan pada penelitian ini, ada kata yang tersedia dalam KBBI akan tetapi tidak tersedia dalam Tesaurus Bahasa Indonesia.

PUSTAKA

- [1] D. Zamzami, F. Rahutomo, and D. Puspitasari, "Aplikasi wordnet indonesia berdasarkan kamus thesaurus bahasa indonesia menggunakan algoritma rule based text parsing," in *Seminar Informatika Aplikatif Polinema*, 2016.
- [2] D. D. Putra, A. Arfan, and R. Manurung, "Building an Indonesian WordNet," in *Proceedings of the 2nd International Malindo Workshop*, 2008, pp. 12–13.
- [3] F. Bond, H. Isahara, S. Fujita, K. Uchimoto, T. Kuribayashi, and K. Kanzaki, "Enhancing the Japanese WordNet," in *Proceedings of the 7th workshop on Asian language resources*. Association for Computational Linguistics, 2009, pp. 1–8.
- [4] S. ElKateb, W. Black, H. Rodríguez, M. Alkhalifa, P. Vossen, A. Pease, and C. Fellbaum, "Building a WordNet for Arabic," in *LREC*, 2006, pp. 29–34.
- [5] O. Bilgin, Ö. Çetinoğlu, and K. Ofazer, "Building a WordNet for Turkish," *Romanian Journal of Information Science and Technology*, vol. 7, no. 1-2, pp. 163–172, 2004.

- [6] G. A. Miller and C. Fellbaum, "Wordnet then and now," *Language Resources and Evaluation*, vol. 41, no. 2, pp. 209–214, 2007.
- [7] R. Richardson and A. F. Smeaton, "Using WordNet in a Knowledge-Based Approach to Information Retrieval," 1995.
- [8] R. Mandala, T. Takenobu, and T. Hozumi, "The use of WordNet in Information Retrieval," *Usage of WordNet in Natural Language Processing Systems*, 1998.
- [9] P. Clark, C. Fellbaum, and J. Hobbs, "Using and Extending WordNet to Support Question-Answering," in *Proceedings of the 4th Global WordNet Conference (GWC'08)*, 2008.
- [10] Y. Liu, P. Scheuermann, X. Li, and X. Zhu, "Using WordNet to Disambiguate Word Senses for Text Classification," in *international conference on computational science*. Springer, 2007, pp. 781–789.
- [11] Gunawan, "Synonym Sets Extraction Based Gloss Acquisition Using Supervised Learning," *Disertation*, 2016.
- [12] H. Hendrik and A. B. Cahyono, "Model WordNet Bahasa Indonesia berbasis Linked Data," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 6, no. 1, pp. 8–14, 2017.
- [13] A. Saputra *et al.*, "Building Synsets for Indonesian Wordnet with Monolingual Lexical Resources," in *2010 International Conference on Asian Language Processing*. IEEE, 2010, pp. 297–300.
- [14] T. Redaksi, "Tesaurus Bahasa Indonesia Pusat Bahasa," *Pusat Bahasa, Departemen Pendidikan Nasional*, 2008.
- [15] E. Setiawan, "Kamus Besar Bahasa Indonesia (KBBI) Online," *KBBI Indones*, 2012.
- [16] Setiawan, Ebta, "Kamus Besar Bahasa Indonesia (KBBI)," *Badan Pengembangan dan Pembinaan Bahasa, Kemdikbud (Pusat Bahasa)*, vol. 2016, 2012.
- [17] L. Sulistyono-Basuki, *Teknik dan Jasa Dokumentasi*. Gramedia Pustaka Utama, 1992.
- [18] A. Setiyoadi, L. Muflikhah, and M. A. Fauzi, "Named Entity Recognition Menggunakan Hidden Markov Model dan Algoritma Viterbi pada Teks Tanaman Obat," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, vol. 2548, p. 964X, 2017.