

PEMBANGUNAN BASIS DATA LEKSIKAL WORDNET BAHASA SUNDA

BUILDING LEXICAL DATABASE SUNDA WORDNET

Novihana Nurani Setiawan

Prodi D3 Manajemen Informatika, Fakultas Ilmu Terapan, Universitas Telkom

novihananns@students.telkomuniversity.ac.id

Abstrak

WordNet adalah suatu sumber informasi rujukan yang banyak dipakai dalam berbagai aplikasi mutakhir yang membantu manusia mengolah dan mencerna informasi, misalnya aplikasi-aplikasi *information retrieval*, *machine translation*, dan *natural language processing (NLP)*. *WordNet* bahasa Sunda adalah kamus bahasa Sunda yang memfokuskan pada makna kata. Satu makna dalam *WordNet* dapat dinyatakan dengan *synset* yaitu kumpulan kata yang merepresentasikan suatu makna. Selain representasi makna, di dalam *WordNet* juga terdapat relasi antar makna seperti sinonim, antonim, hipernim, hiponim, holonim, meronim. Proyek Akhir ini bertujuan untuk membangun *WordNet* bahasa Sunda berdasarkan konsep *Princeton WordNet (PWN)*. Data-data yang digunakan antara lain bersumber dari *WordNet* bahasa Indonesia, Kamus Umum Bahasa Sunda (KUBS), dan Kamus Besar Bahasa Indonesia (KBBI). Metode yang digunakan adalah metode *merge*, dengan langkah pertama yang dilakukan adalah membandingkan dua bahasa untuk mencari kesamaan dan perbedaannya, unsur penentu yang dibandingkan berupa bentuk leksikal bahasa Sunda dengan bahasa Indonesia. Proses ini dilakukan dengan menggunakan aplikasi *web-based* dan melibatkan sejumlah pengguna untuk berpartisipasi dengan cara menentukan sinonim, antonim, hipernim-hiponim, dan holonim-meronim dari *sinset* yang telah di sediakan aplikasi *WordNet* bahasa Sunda.

Kata kunci : *WordNet*, Sinonim set, Relasi Leksikal, Relasi Semantik.

Abstract

WordNet is a source of reference information that is widely used in a variety of cutting-edge applications that help people process and digest the information, such as information retrieval applications, machine translation and natural language processing (NLP). *WordNet* Sundanese is a Sundanese dictionary, which focuses on the meaning of the word. One meaning in *WordNet*, can be expressed with *synset*, a collection of words that represent a meaning. In addition to the representation of meaning, in *WordNet* there are also relationships between meanings, such as synonyms, antonyms, hipernim, hyponymy and hypernymy, holonim, meronim. This final project aims to build *WordNet* Sundanese based on the concept of *Princeton WordNet (PWN)*. The data used include *WordNet* bahasa Indonesia, General Dictionary Sundanese (KUBS), and Kamus Besar Bahasa Indonesia (KBBI). The method is used mainly merge method, the first step is to compare the two languages to look for similarities and differences, the decisive element that compared a form of lexical Sundanese with Indonesian. This process is done using a web-based application and involve a number of users, to participate by identifying synonyms, antonyms, hipernim-hyponymy and hypernymy, and holonim-meronim of *sinset*, which has been provided *WordNet* Sundanese application.

Keywords: *WordNet*, *Synonym sets*, *Lexical Relationships*, *Semantics Relation*.

1. Pendahuluan

Setiap kata dalam bahasa Sunda dapat memiliki satu atau lebih makna, dan beberapa kata yang berbeda dapat memiliki makna sama yang disebut dengan sinonim. Sebuah kata pun bisa mempunyai lawan kata yang disebut dengan antonim, selain itu sebuah kata juga dapat memiliki keterhubungan antarkata dengan kata lainnya seperti hipernim, hiponim, holonim, meronim, dan lain-lain. Hipernim merupakan hubungan antar kata yang merupakan bentuk kata yang lebih umum, sedangkan hiponim adalah hubungan antar kata yang menunjukkan bentuk kata yang lebih spesifik. Hubungan antar kata lainnya seperti meronim yang merupakan hubungan antar kata yang menunjukkan kata yang merupakan bagian dari anggota kata yang lain, sedangkan holonim merupakan kata yang menjadi bagian dari keseluruhan dari kata lainnya.

Dalam penyajian kata-kata seperti dalam kamus, keterhubungan antar kata tersebut tidak disebutkan. Hal ini menimbulkan kesulitan dalam mencari keterhubungan antar kata. Kamus Umum Bahasa Sunda belum menunjukkan keterhubungan antar kata tersebut. Oleh karena itu, dibutuhkan sebuah sistem yang dapat menanggulangi kekurangan tersebut. Salah satunya adalah mengembangkan aplikasi *WordNet* yang merupakan sebuah basis data leksikal elektronik. Dalam *WordNet* bahasa Sunda, kata dalam bahasa Sunda dikelompokkan berdasarkan kesamaan makna menjadi himpunan sinonim (*sinset*). Sebuah *sinset* dapat dipandang sebagai representasi sebuah konsep semantik.

Perbedaan antara *WordNet* dan kamus bahasa pada umumnya terletak pada fokus dari masing-masing metode. Kamus bahasa secara umum memfokuskan pada kata itu sendiri, sedangkan *WordNet* berfokus pada makna kata. *WordNet*, merupakan kamus semantik yang semula dibuat dalam versi bahasa Inggris dan menjadi salah satu sumber informasi rujukan yang banyak di pakai dalam berbagai aplikasi mutakhir. Pembangunan *Wordnet* bahasa Sunda bertujuan untuk membangun suatu aplikasi yang menampung dan menampilkan keterhubungan kata, selain itu pengembangan aplikasi ini diharapkan dapat membuka peluang diterapkannya berbagai jenis ilmu dan teknologi terhadap sumber informasi bahasa Sunda.

2. Dasar Teori/Material dan Metodologi/perancangan

2.1 WordNet

Pada tahun 1985, beberapa psikolog dan linguist dari *Princeton University* mengembangkan basis data leksikal yang bertujuan sebagai sarana untuk membantu pencarian secara konseptual dari pada secara alfabetis ke dalam kamus. Hasil dari pengembangan ini adalah sebuah kamus dengan pendekatan psikolinguistik yang disebut sebagai *WordNet* [3]. Perbedaan antara *Princeton Univertisy WordNet* dengan kamus pada umumnya adalah pengaturan informasi leksikal berdasarkan makna kata bukan berdasarkan kata itu sendiri.

Pada awal pengembangan *Princeton WordNet*, kata dan makna kata dapat diilustrasikan dengan sebuah matriks dengan makna kata sebagai baris (M_1, M_2, \dots, M_m) dan kata sebagai kolom (F_1, F_2, \dots, F_n) (Lihat Tabel 1) [2]. Dan Entri $E_{1,1}$ menyatakan bahwa kata F_1 dapat digunakan untuk menyatakan makna M_1 . Jika dalam satu kolom yang sama terdapat minimal dua buah makna yang dinyatakan oleh satu kata, maka kata tersebut bersifat polisemi (Contoh. $E_{1,2}, E_{2,2}$). Jika dalam satu baris yang sama terdapat minimal dua buah kata, maka kata-kata tersebut disebut sebagai sinonim (Contoh. $E_{1,1}, E_{1,2}$).

Table 1 Matriks Leksikal

Makna Kata	Kata				
	F_1	F_2	F_3	...	F_n
M_1	$E_{1,1}$	$E_{1,2}$			
M_2		$E_{2,2}$			
M_3			$E_{3,3}$		
...				...	
M_n					$E_{n,n}$

Pemetaan antar kata dengan makna kata yang terjadi pada Tabel 1 merupakan pemetaan *many-to-many*, dimana satu makna dapat dinyatakan dengan satu atau beberapa kata dan sebaliknya, satu kata dapat menyatakan satu atau beberapa makna. Maka M_1 dapat direpresentasikan dengan mendaftarkan semua kata yang digunakan ke dalam sebuah himpunan $\{F_1, F_2\}$. Elemen-elemen yang termasuk dalam himpunan tersebut disebut sebagai sinonim dan himpunan tersebut disebut sebagai *synset* (*synonym set*).

Secara lengkap, satu makna dalam *Princeton WordNet* dapat dinyatakan dengan *synset*, kelas kata dari *synset*, definisi dari makna tersebut dan contoh penggunaan dalam kalimat. Sebagai contoh, $\{arrangement, organisation, organization, system\}$ merupakan sebuah *synset* yang memiliki kelas kata sebagai kata benda (*noun*) dan menyatakan makna *an organized structure for arranging or classifying*. Ada beberapa contoh penggunaan *synset* tersebut dalam kalimat, yaitu *he changed the arrangement of the topic; the facts were familiar but it was in the organization of them that he was original; he tried to understand their system of classification*.

2.2 Relasi Semantik

Beberapa relasi semantik yang terdapat di dalam *Princeton WordNet* antara lain:

1. Hponim, merupakan hubungan antara makna spesifik dan makna generik atau antar anggota taksonomi [4]. Sebuah makna 'X' dikatakan sebagai hponim dari makna 'Y' apabila makna 'X' merupakan jenis dari makna 'Y' [6]. Sebagai contoh, {mawar, melati, anggrek} merupakan hponim dari {bunga}.
2. Hipernim, merupakan relasi semantik yang merupakan kebalikan dari relasi hponim. Sebuah makna 'Y' dikatakan sebagai hipernim dari makna 'X' apabila makna 'X' merupakan jenis dari makna 'Y' [4]. Dari contoh sebelumnya dapat dikatakan bahwa {bunga} merupakan hipernim dari {mawar, melati, dan anggrek}.
3. Meronim, merupakan relasi yang menyatakan sebuah makna sebagai bagian dari makna yang lain. Sebuah makna 'X' dikatakan sebagai meronim dari makna 'Y' apabila makna 'X' merupakan bagian dari makna 'Y' [6]. Di dalam *Princeton WordNet*, meronim dapat di bagi menjadi 3, yaitu :
 - a. *Part* meronim. Sebagai contoh, {kuku} merupakan meronim dari {jari}.
 - b. *Member* meronim. Sebagai contoh, {profesor} merupakan meronim dari {fakultas}.
 - c. *Substance* meronim. Sebagai contoh, {oksigen} merupakan meronim dari {udara}.
4. Holonim merupakan relasi semantik yang merupakan kebalikan dari meronim. Sebuah makna 'Y' dikatakan sebagai holonim dari makna 'X' apabila makna 'X' merupakan bagian dari makna 'Y' [4]. Di dalam *Princeton WordNet*, holonim dibagi menjadi 3, yaitu :
 - a. *Part* holonim. Sebagai contoh, {jari} merupakan holonim dari {kuku}
 - b. *Member* holonim. Sebagai contoh, {fakultas} merupakan holonim dari {profesor}.
 - c. *Substance* holonim. Sebagai contoh, {udara} merupakan holonim dari {oksigen}.

Relasi semantik seperti hponim, hipernim, meronim, dan holonim merupakan relasi yang bersifat transitif [3]. Sebagai contoh {mawar, melati, anggrek} merupakan hponim dari {bunga} dan {bunga} merupakan hponim dari {tanaman} maka dapat disimpulkan bahwa {mawar, melati, anggrek} merupakan hponim dari {tanaman}.

2.3 Relasi Leksikal

Relasi leksikal merupakan relasi yang dibentuk antar kata. Beberapa contoh relasi leksikal diantaranya adalah antonim, *pertainym*, *participle*, dll [4].

1. Antonim, relasi antonim adalah kata-kata yang memiliki pertalian makna bertentangan secara penuh atau secara sebagian [5] .

2.4 Metode Komparatif

Metode yang digunakan dalam pembangunan *WordNet* bahasa Sunda adalah metode komparatif (*Merge*). Dengan menggunakan sumber-sumber data yang berasal dari Kamus Umum Bahasa Sunda (KUBS) dan Kamus Besar Bahasa Indonesia. *Dutch WordNet* merupakan salah satu *WordNet* yang dikembangkan dengan metode ini [1]. Secara garis besar, ada dua tahap yang akan dilakukan dalam pembangunan *WordNet* bahasa Sunda, yaitu

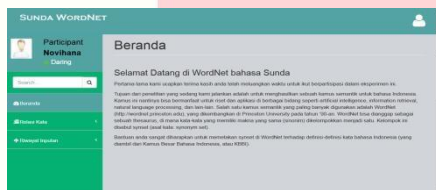
1. **Tahap 1** : Pemetaan makna dan kata yang terdapat dalam *Princeton WordNet* kedalam makna dan kata dalam bahasa Sunda dengan bantuan Kamus Umum Bahasa Sunda (KUBS) dan kamus dwi-bahasa.
2. **Tahap 2** : Pengolahan hasil pemetaan untuk menghasilkan basis data *WordNet* bahasa Sunda.

3.2 Implementasi



Gambar 6 Implementasi Halaman Login

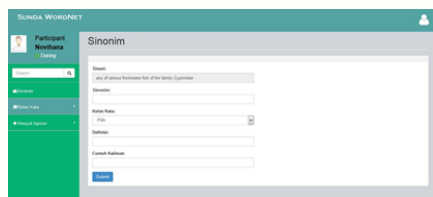
3.2.1 Implementasi Antarmuka Partisipan



Gambar 7 Halaman Beranda Partisipan Gambar



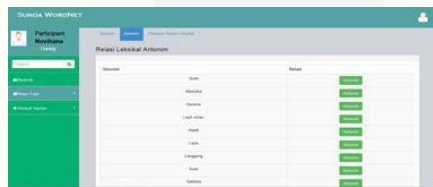
8 Halaman Relasi Leksikal Sinonim



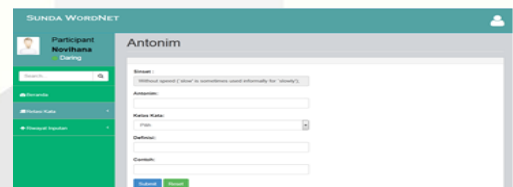
Gambar 9 Halaman Input Sinonim



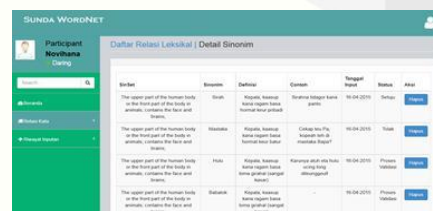
Gambar 10 Halaman Detail Sinonim



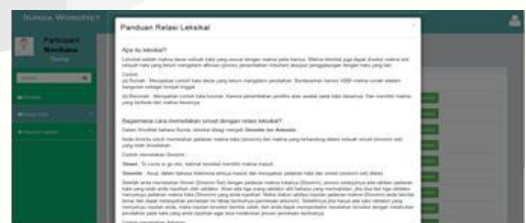
Gambar 11 Halaman Relasi Leksikal Antonim



Gambar 12 Halaman Input Antonim



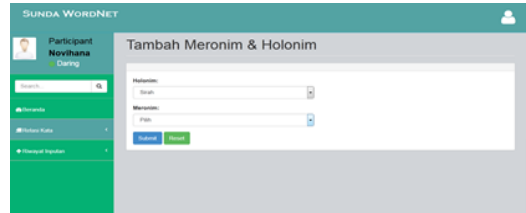
Gambar 13 Halaman Detail Antonim



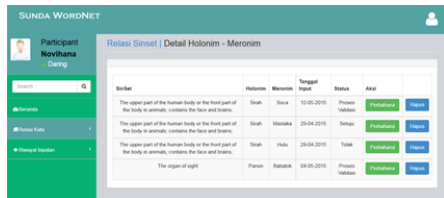
Gambar 14 Halaman Panduan Relasi Leksikal



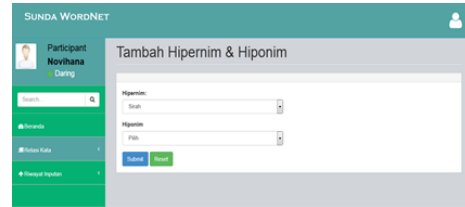
Gambar 15 Halaman Relasi Semantik



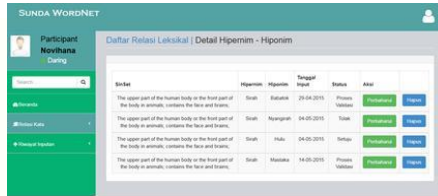
Gambar 16 Halaman Input Holonim & Meronim



Gambar 17 Halaman Detail Holonim & Meronim



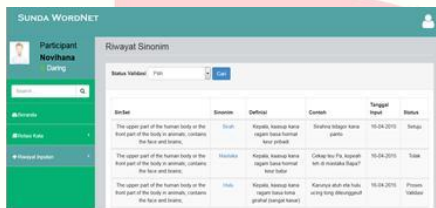
Gambar 18 Halaman Input Hipernim & Hiponim



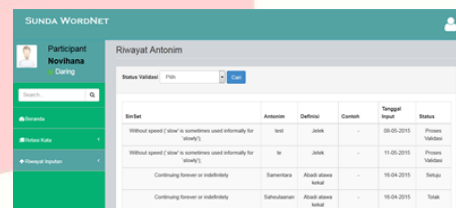
Gambar 19 Halaman Detail Hipernim & Hiponim



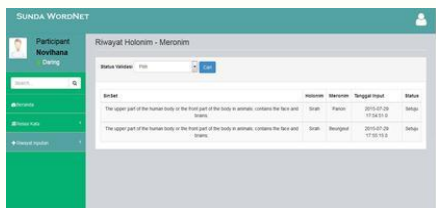
Gambar 20 Halaman Panduan Relasi Semantik



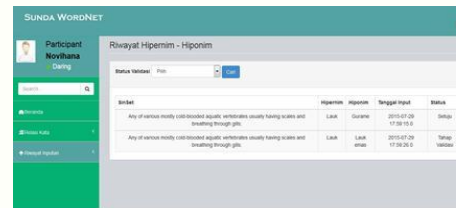
Gambar 21 Halaman Riwayat Sinonim



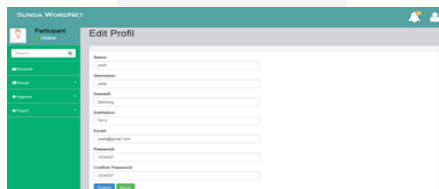
Gambar 22 Halaman Riwayat Antonim



Gambar 23 Halaman Riwayat Holonim & Meronim

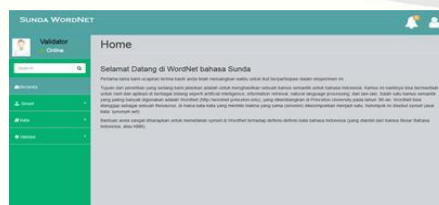


Gambar 24 Halaman Riwayat Hipernim & Hiponim

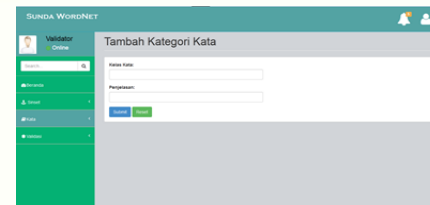


Gambar 25 Halaman Edit Profil

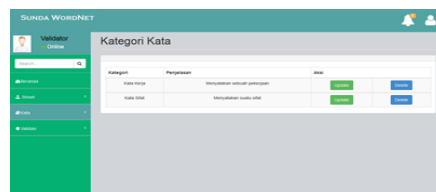
3.2.2 Implementasi Antarmuka Validator



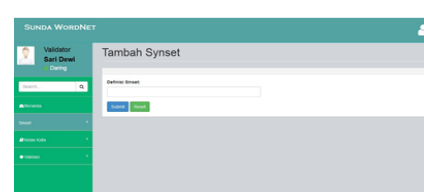
Gambar 26 Halaman Beranda Validator



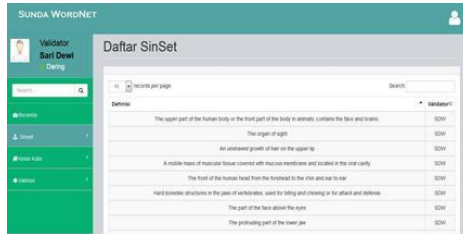
Gambar 27 Halaman Input Kategori Kata



Gambar 28 Halaman Daftar Kategori Kata



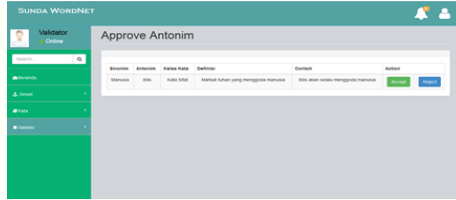
Gambar 29 Halaman Input Synset



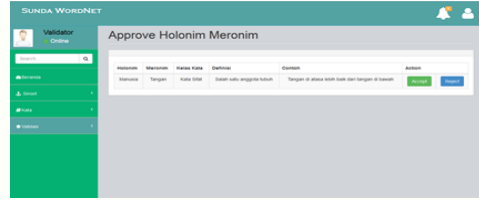
Gambar 30 Halaman Daftar Synset Gambar



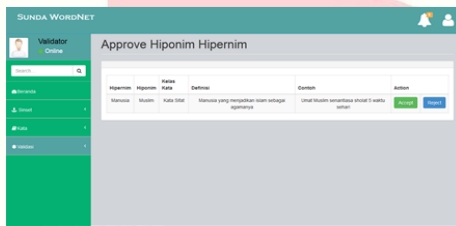
31 Halaman Validasi Sinonim



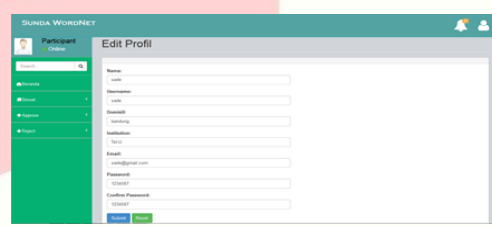
Gambar 32 Halaman Validasi Antonim



Gambar 33 Halaman Validasi Holonim & Meronim



Gambar 34 Halaman Validasi Hipernim & Hiponim



Gambar 35 Halaman Edit Profil

3.2.3 Implementasi Antarmuka Admin



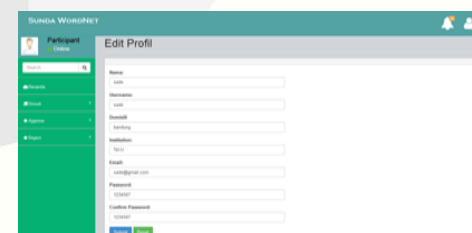
Gambar 36 Halaman Beranda Admin Gambar



37 Halaman Input Data Validator



Gambar 38 Halaman Daftar User



Gambar 39 Halaman Edit Profil

4. Kesimpulan

Adapun kesimpulan dari proyek akhir ini adalah sebagai berikut :

1. Aplikasi *WordNet* bahasa Sunda merupakan aplikasi yang menyimpan himpunan sinonim (sinset) dan keterhubungan antar kata dalam bahasa Sunda.
2. Aplikasi *WordNet* bahasa Sunda memiliki fitur pencarian kata, yang dapat membantu menampilkan keterhubungan antar kata.
3. Aplikasi *WordNet* bahasa Sunda memiliki fitur validasi data, sehingga data yang ada pada *WordNet* bahasa Sunda merupakan data yang valid.

Daftar Pustaka:

- [1] Bloksma, L., Boersma, P., & Vossen, P. 1999 *The Deutch WordNet*. Amsterdam: Vrije Unibersitei.
- [2] Fellbaum, C. 1998 *A Semantic Network of English: The Mother of All Computers and the Humanities* 32, 209-220.
- [3] Miller, G.A., Beckwith, R., & Fellbaum, C. 1993 *Introduction to WordNet: An On-line lexical Database*.
- [4] Alwi, Hasan dkk. 2002. *Kamus Besar Bahasa Indonesia Edisi ketiga*. Jakarta. Balai Pustaka.
- [5] Kusri. 2007. *Strategi Perancangan dan Pengelolaan Basis Data*. Yogyakarta. Amikom .
- [6] WordNet 3.0 Reference Manual. (2006). Retrieved June 20, 2008, from <http://wordnet.proncton.edu/doc>

