DETECTION OF RELATIONS ON THE MEENING OF WORDNET IN INDONESIA LANGUAGE

Muhamad Iffandi Pribadi[1], Ken Kinanti P S.Kom, M.T[2]

[1,2] Teknik Informatika – Universitas Komputer Indonesia
Jalan Dipatiukur 114-116 Bandung 40132
E-mail : iffandipribadi@gmail.com[1], kinantiken@gmail.com[2]

ABSTRACT

Wordnet is a network between Words that inspired from human lexical memory. It is also described as a place to create an online lexical Database to help lexicographer, researcher, or a language activist to search information or the semantic relation of a lemma. Wordnet also can be applied on applications that help human to processing and understanding information like in The Informal Retrieval field, Machine Translation Field, and Natural Language Processing Field.

The development of this Indonesian Wordnet is using an Expand Approach with Automatic translation method. The resource of the data was taken from Princeton Wordnet and using MRD from Cambridge Dictionary.Extraction process was done on the first step so the structured PWN Database's data can be obtained. The trasnlation process from English to Indonesian was done using MRD from Cambridge Dictionary in The Second Step.

In the first stage, the results of Unique Strings extraction as much as 100% of the total initial data 155467 and Synset as much as 100% of the total initial data as much as 117791. In the second phase of the translation process obtained Unique Strings 98.78%, 98.48% for antonym, 93.40% for hipernim / Hyponim, and 95.33% for meronym / holonym. That's because a lot of local entries or terms are in English English that do not exist in the Indonesian language.

KeyWords : Wordnet, Indonesian Wordnet, expand approach, Automatic Translation

.

1. INTRODUCTION

The development of WordNet has now been tried in many countries such as Arabic, Spanish, French, Dutch etc. Although Wordnet First was created manually, further development was done with automated and semi-automated techniques to create Multilingual WordNet. The method is generally divided into two ways: Merge Approach and Expand approach. Merge Approach is a method used by Princeston University to create WordNet, in addition to spending a lot of time, merge approach is also very expensive to build, because it must involve lexicographer to create syns. Then the next approach is Expand Approach, in contrast to merge approach, expand approach translate synset existing in Princeston WordNet (PWN) to target Language and take all relation in PWN. The Expand approach requires manual validation so that the resulting information is not ambiguous.

In previous research on wordnet development of indonesian data generated from category noun, verb, adj, and adverb result obtained for Synset is equal to 40774 from total 117791 Synset and Unique Strings generated 23964 from 155467. At translation process of noun category that have big impact to The percentage of test results is only 15.4% for Unique Strings. The problem in previous research is the incomplete translation of translation result using MRD Cambridge Dictionary because the word form of noun mostly is a regional noun term in English. Many Medical, Chemical, and Term terms for other specialized dictionaries that can not be translated and compound words in English are very few of its Indonesian translations [1]

Based on the above problem, this research will develop previous research by building Indonesian wordnet application by completing the elements of inter-meaning relation such as antonyms, hyponims, holonyim-meronim and improve the result of Unique Strings and Synset by using expand approach. By using Oxford Dictionary, KBBI and Thesaurus, this research uses Oxford Dictionary resource because Oxford Dictionary one of the most complete dictionary and translation data can be for Crawling from Oxford English Dictionary website. Translation process by taking the good word from index.pos and data.pos to then be converted into target language that is Indonesian as target resource of Language, currently Oxford Dictionary has released the eighth edition with currently number 600,500 entries that become reference for Making wordnet Bahasa Indonesia. With the help of MRD Dualingual English-Indonesian and Indonesian English or

machine translation for translation from PWN to Bahasa Indonesia and manual validation is done to avoid ambiguity of MRD and machine translation so that the resulting data can be qualified.

2. RESEARCH CONTENTS

Wordnet development using Expand Approach with Atutomatic translation.

2.1 Problem analysis

The problem drawn from the background discussed earlier is to develop from previous research. Due to the lack of structure regarding Wordnet Database, the relation between open source meaning. And the absence of semantic network results from Lexicographer, for that relationship between the meaning of the Indonesian language has not been formed and available. For now the development of Wordnet Indonesian has not been re-done and the absence of resources that can be used and can be developed with open source properties. In previous research on wordnet development of indonesian language the resulting data is from category noun, verb, adj, and adverb test result obtained from previous development can be seen in following table :

1. Table Results Testing

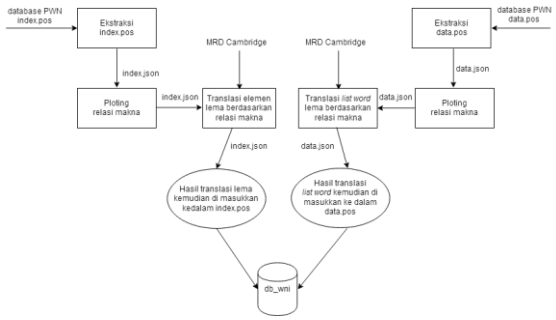
Category	Preliminary data		Extraction Results		Translated Results	
	Unique Strings	***Synset***	***Unique Strings***	***Synset***	***Unique Strings***	***Synset***
Noun	117953	82192	117953	82192	11944	20992
Verb	11540	13789	11540	13789	5283	10290
Adjective	21499	18185	21499	18185	4755	7366
Adverb	4475	3625	4475	3625	1982	2126
TOTAL	**155467**	**117791**	**155467**	**117791**	**23964**	**40774**

Synset at get result of 40774 from total 117791 Synset and Unique Strings generated 23964 from 155467 Unique Strings. In the process of translation noun categories that have a major impact to the percentage of test results is only obtained about 15.4% for Unique Strings. The problem in previous research is that there is no translation of translation result using MRD Cambridge Dictionary because the word form of noun mostly is a regional noun term in English. Many Medical, Chemical, and Term terms for other specialized dictionaries that can not be translated and compound words in English have very few Indonesian translations. Wordnet's need for information is still unstructured due to the development of a new dictionary so that it takes a resource to process the Indonesian language.

2.2 Solution Analysis

Solution analysis is a stage in which the process is to identify any problems and needs needed to make the experiment in another way. The solution given in this research is to use the Expand Approach Method with Automatic Translation by translating any entries or words contained in Wordnet Princeton Database (PWN) by extracting Database from Princeton Wordnet and retrieve any existing semantic relation on PWN. The first step is the process of cleaning character that is not data, such as license agreement, number version and copyright notice that existed at the beginning of each database. Because the format of Database PWN in the form of data and index and data structure is different for that will be done two different process extract Database based on data and index. After the extraction phase successfully translated using Machine Readable Dictionary (MRD) from the Cambridge Dictionary on every entry in the PWN Database. The next one is saved into the new data. For more details can be seen in Figure 1.

In the Analysis of solutions for the Wordnet development system the Indonesian language using MRD has the extract stages of the PWN Database, translation entries, and format conversion.



Picture 1 Figure Overview of System Architecture

Figure 3.1 Overview of the system architecture described above as follows.

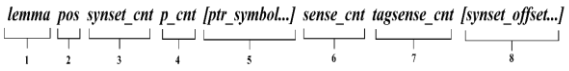
1. In the early stages of User enter Databases PWN index.pos ASCII-formatted into the system and then extraction of data structures.
2. Next ploting each entry on the index.pos based on the relation of meaning.
3. The next stage is done the translation process on the first element is the entry. Translations using MRD from Oxford Dictionary
4. The translated entry is then sorted and stored into the Database.
5. In this process the input data is Database PWN data.pos that is formatted ASCII into the system and then extraction and retrieval of data structures, the output format is .json
6. Next ploting each entry on the data.pos based on the relation of meaning.
7. At this stage Wordslist is taken on every line and translated from English to Indonesia with the help of MRD.
8. After successful translation then take the existing data structure on the PWN and enter the translation into the structure of the data taken and stored into the Database.

2.3 Input Data Analysis

Analysis of data input on the system that explains the data process of Database from PWN and data from Cambridge Dictionary. The PWN database itself has a special format by grouping data with data.pos and index.pos where this heading is noun, verb, adjective, and adverb. The PWN Database format uses ASCII encoding so it is easy to read by humans and machines.

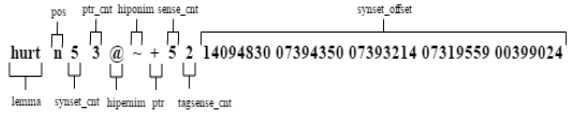
2.3.1 Index.pos

The file index consists of an alphabetical word list with the Word class of each Database. On each line the word list has an element called Synset_offset connected to data.pos, the relation between meanings represented by the data structure [ptr_symbol ...], each Synset contains the word. The word / index on index.pos has a lower case format. The index.pos of each Database consists of copyright, version number, license agreement at the beginning of the line. Here is the data structure that exists on index.pos.



Picture 2 Data Input *index.pos*

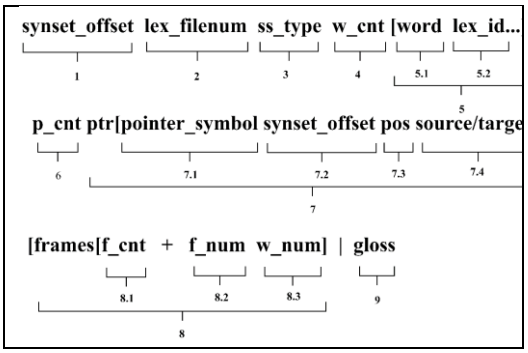
For more details here is an example of input data index.pos.



Picture 3 Sample Data Input Index.pos

2.3.2 *Data.pos*

The data.pos file has information containing Synset members from index.pos. Each data.pos file starts with copyright notice, version number and license agreement. Each data row has information that has been compiled by Lexicographer for Synset. Each data information is initiated by an 8 byte offset or address of a Synset.



Picture 1. 1 Data Input *Data.pos*

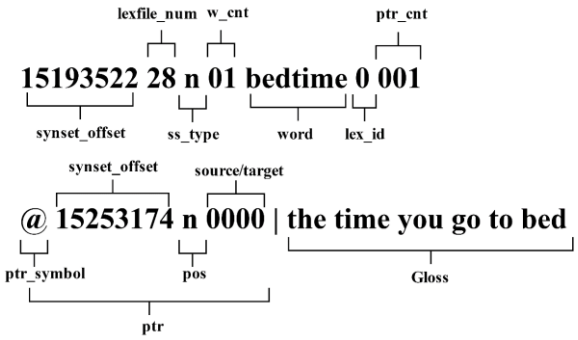
Following explanation of the data structure *data.pos.*

Tabel 3. 1 Tabel Elemen *Data.pos*

No.	Element	Explanation
1	*Synset_offset*	8 byte decimal integer containing address for Synset information
2	*Lex_filenum*	2 digit decimal integer corresponds to data from a Lexiconographer containing Synset.
3	*Ss_type*	The word class for each entry, consisting of n (noun), v (Verb), a (adjective), r (adverb).
4	*w_cnt*	2 integer hexadecimal digits that contain the number of words in Synset
5.1	*Word*	Word with the lower case ASCII format. Spaces are replaced with underscore (_). Particularly in data.adj followed by syntactic marker if the data has been specifically defined in the Lexicon data.
5.2	*Lex_id*	One integer hexadecimal digit in word. Refers to the lexicographer file. Usually lex_id starts from the number 0. 0 including the default status and has no Lexicon data.
6	*P_cnt*	three digit decimal integer contains about the number of Pointer on a Synset. P_cnt is worth 000 then Synset has no Pointer
7	*Ptr*	*Pointer on a Synset, has a structure: Pointer_symbol offset post source / target*
7.1	*Ptr_symbol*	List Pointer available on sysnet. Pointer is a semantic relation that exists on every Synset.
7.2	*Synset_offset*	8 byte decimal integer containing address for Synset information
7.3	*Pos*	The word class for each entry, consisting of n (nount), v (Verb), a (adjective), r (adverb)

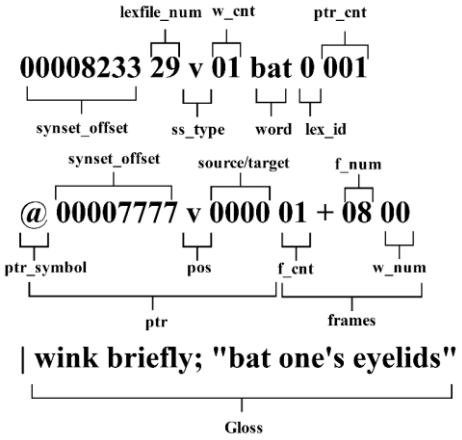
7.4	*Source/Target*	*Source / Target is lexical and semantic Pointer. 4 digits of character, the first 2 digits are the first 2-digit integer hexadecimal indicating the number of words in Synset. And the last 2 digits indicate the number of words on the target Synset.*
8	*Frames*	*Frames only exist in data.verb, has a form* *F_cnt + f_num w_num*
8.1	*F_cnt*	A two-digit decimal integer that identifies the number of frames present on a line.
8.2	*F_num*	Two integer hexadecimal digits that denotes the number of a frames
8.3	*W_num*	Two integer hexadecimal digits that indicate the number of words containing frames,
9	Gloss	Each Synset has a gloss or definition, separated by a pipe character "|". In the gloss element there are definitions and sentence examples

Berikut adalah contoh data masukan *data.pos.*



Picture 3. 4 Sample Data Input *Data.pos*

There is little difference to the data.verb that is the frames element, Here is a sample data.verb for frames.



Picture 3. 5 Sample Data Input *Data.verb*

2.4 Process Analysis

The process analysis explains the systematic approach in identifying each problem and goal in designing the Indonesian Wordnet. The process analysis performed is extraction and translation as the main process and data conversion. Because the data structure on the data.pos and index.pos different then done different extract process for it done 2 different process.

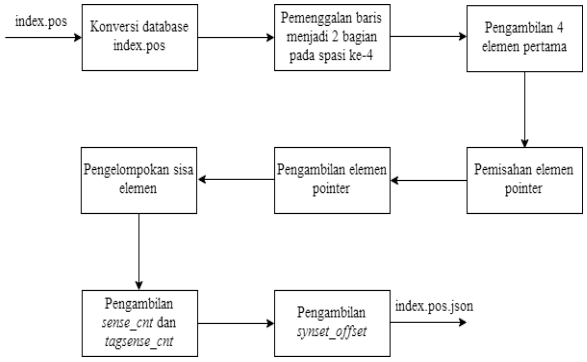
2.4.1 Extraction Index.pos

This stage focuses on retrieving the data structure on index.pos to get the structure consisting of lemma, post, Synset_cnt, p_cnt [ptr_simbol ...] sense_cnt, tagsense_cnt and Synset_offset [Synset_offset ...]. The input data comes from Princeton Wordnet (PWN) which is in ASCII format. PWN itself splits its Database into four sections each of which are grouped by word class ie index.noun, index.verb, index.adverb, and index.adjective Examples of input data taken from PWN are as follows.

Table 3. 2 Data Index.pos

hurt n 5 3 @ ~ + 5 2 14094830 07394350 07393214 07319559 00399024 laugh n 3 4 @ ~ %p + 3 2 07029036 06785996 06687271

The above stages can be seen in the following figure.



Picture 6 Database Index.pos Schema Extraction Scheme from PWN

Table 3. 3 Data *Index.pos*

```
commerce n 3 6 @ ~ %p = + - 3 1 01092370
08146250 07152330

commando n 2 3 @ ~ ; 2 1 09961754 08421330
```

The above table of the PWN database is the source of data to extract and retrieve the data structure. The purpose of the created system will take each part of the structure and insert it into the table in the database

Here are the steps performed in database index.pos extraction.

2.5 Extraction *Data.pos*

After taking the database index.pos the next stage is taking data.pos which contains the data from the index. This stage focuses on retrieving data structures on the data.pos to get the structure of Synset_offset, lex_filenum, ss_type, w_cnt, word, lex_id, [word lex_id ...], p_cnt, [ptr ...], [frames ...] , | Gloss,. The input data comes from Princeton Wordnet (PWN) in ASCII format, just as the PWN index divides its database by post or word class ie data.noun, data.verb, data.adv, and data.adj. Examples of input data taken from PWN are as follows.

Table 1 Sample Data Input *Data.pos*

```
08497858 14 n 01 blue 0 002 @
08024893 n 0000 ~ 08411492 n 0000
| any organization or party whose
uniforms or badges are blue; "the
Union army was a vast blue"

08497673 14 n 01 vote 0 000 | a
body of voters who have the same
interests; "he failed to get the
Black vote"
```

Data masukan pada tabel di atas adalah *data.pos* From the PWN database which becomes the data source to extract and retrieve the data structure. The goal is the same as the index to retrieve each element. For the data structure is more complex and more difficult than index.pos for each element is not all the same there is every line. Some differences on each such data.

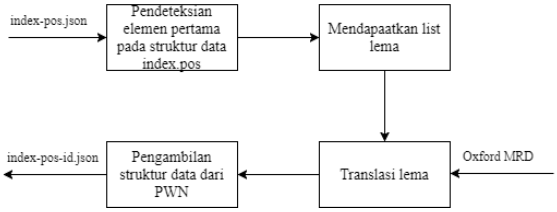
1. *[frames...]*, whose position exists before the gloss exists only in the data.verb, and does not exist in other data.
2. For some cases there are lines that do not have a word definition
3. For some cases there is a line that has no word definition.
4. Some lines have no sentence examples.
5. Some lines do not have Pointer.

2.6 Translating Index.pos and Data.pos

After the extraction process is done the translation of each Synset on the database that has been extracted. The input data is extracted index.pos and data.pos with .json format. The translation process is translated by Machine readable Dictionary (MRD) from Cambridge Dictionary. The reason for using the Cambridge Dictionary MRD is because of the amount of data that can be retrieved from the website. In this process the system will check whether the translated list will be translated or not, if obtained terjemahanya then the first element will be updated and filled with the results of the Indonesian translation, whereas if not found the translation then the entry is filled with empty fields. It is useful for validating how many entries have been successfully translated.

2.6.1 Translating.pos

In the process of translating the index.pos element taken is the first element of the data structure that is the element of the entry. The entry element is a Unique String on the Wordnet Princeton database that is used to search within a Synset member. The process of translation is done as follows.

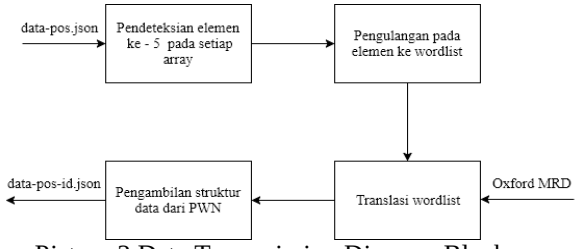


Picture 1 Block Translation Process Diagram *Index.pos*

As we see the structure of the data taken is the first element that is the entry, in the example below there is a list of unique strings taken are commerce and commando. The system translates and then generates translations for the target language and is re-entered into the data structure of Princeton Wordnet.

2.6.2 Translations Data.pos

In Process of data.pos data is taken from data.pos extraction that is element 5 that is Wordlist. For word list Because the number of words in each line can be more than one then repeated to do the translation after that the process of taking the data structure from the extraction. For more details can be seen in the block diagram below.



Picture 2 Data Transmission Diagram Block.*pos*

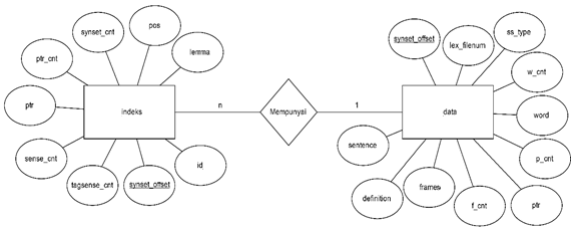
In the 5th element of each array, it is repeated if the word has more than one. Then translated from the MRD Cambridge Dictionary on the word list, then the word list is inserted into the data structure for each row.

2.7 Database Analysis

This analysis will be described with ERD and described in the data dictionary ERD.

2.7.1 Entity Relationship Diagram

The system to be built will use the ERD scheme below.



Gambar 3 *Entity Relationship Diagram* (ERD)

2.7.2 Kamus Data *ERD*

Explanation on the picture Entity relationship The diagram above using the ERD data dictionary, for more details can be seen below.

Table 2 Data Dictionary ERD

Entity	Attribute
t_indeks	*id_indeks, lema,pos,Synset_cnt,ptr_cnt, ptr_sense_cnt, tagsense_cnt, offset.*
t_data	*Synset_offset, lex_filenum, ss_type, w_cnt, word, p_cnt, ptr, f_cnt, frames , definition, sentence.*

2.8 Non-Functional Needs Analysis

Non-functional requirement analysis is done to find out the requirement of specification for system which includes Hardware and software. This analysis is needed to make it easier to analyze a system.

2.8.1 Analysis of Hardware Systems Development Needs

Hardware requirement (hardware) for this development is the minimum specification to run this system, the constraints to be faced is the process will take a long time. For minimum hardware specifications are as follows.

Table 3 Hardware Minimum Specifications

Hardware	Minimum Requirements
Prosesor	Intel Pentium IV 3,0 GHz atau AMD Athlon X4
RAM	1Gb

2.8.2 System Requirement Analysis of System Development

Software requirements analysis (Software) for the construction of this system is as follows.

Table 4 Minimum Software Specification

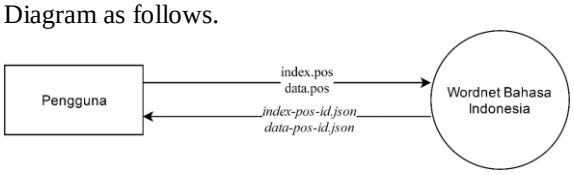
Software	Minimum Version of the Software
Sistem Operasi	Windows XP
Web Browser	Google Chrome,mozilla,opera,safari

2.9 Functional Needs Analysis

The purpose of this analysis is to analyze the process to be applied in the system to be built using Context Diagram, DFD Creation, and Process Specification.

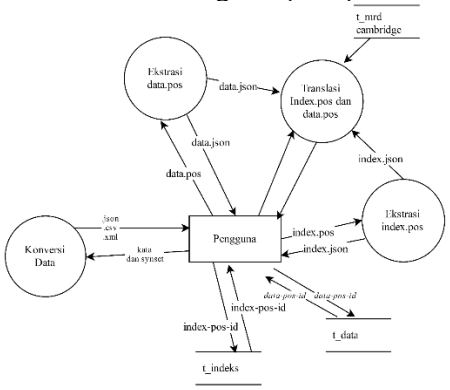
2.9.1 Context Diagram

Context diagram contained in Wordnet Indonesian built can be seen in the picture Context Diagram as follows.



Picture 4 Context Diagram

2.9.2 Data Flow Diagram (DFD) Level 1



Gambar 5 Data Flow Diagram (DFD) Level 1

2.9.3 Data Flow Diagram (DFD) Level 2 Index.pos and Data.pos

The data that has been entered into the retrieval of index.pos and data.pos which is described in the form of DFD is the flow of data entry to become the resulting data. Starting from database extraction from index.pos and data.pos. The result of retrieving this database is stored as .json which will be processed for

the next step in the form of translation of each entry. The input data has its own extension in the form of a post or word class that is in ASCII format, the data to be generated is .json. Here is the DFD from the index.pos and data.pos database retrieval.

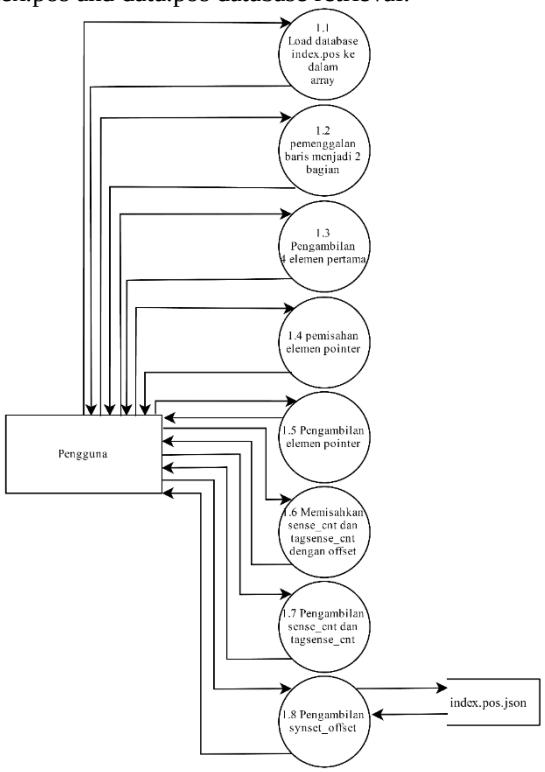
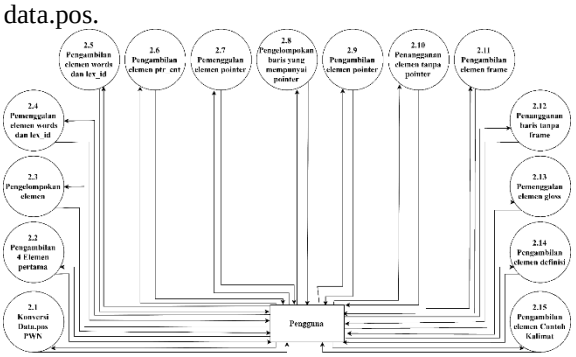


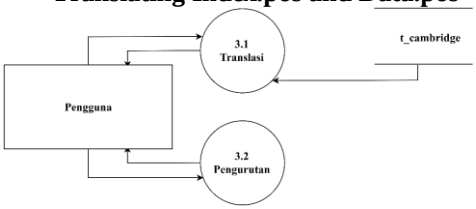
Table 6 DFD Level 2 Extraction Process *Index.pos*

Here is a DFD Level 2 extraction process for data.pos.



Gambar 7 DFD Level 2 Ekstraksi Data.pos

2.9.4 Data Flow Diagram (DFD) Level 2 Translating Index.pos and Data.pos



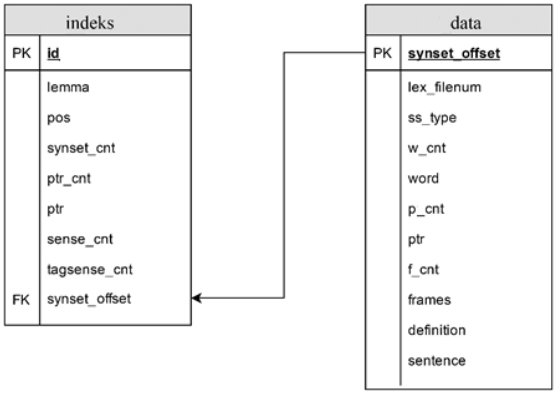
Picture 8 *DFD* Level 2 Translating index.pos and data.*pos*

2.10 Database Design

The design of the database determines the content and arrangement of data needed to support various system design, its purpose to meet the information that contains the needs in Wordnet development of the Indonesian language.

2.10.1 Relation Scheme.

Skema relasi memberikan hubungan antar tabel secara lengkap, termasuk di dalamnya terdapat atribut. Hal ini diperlukan untuk menyimpan informasi mengenai objek-objek yang direpresentasikan dalam sebuah basis data. Tabel yang dibangun untuk Wordnet bahasa Indonesia mengambil struktur data dari *database Princeton Wordnet* (PWN). Sehingga terdapat dua tabel yaitu tabel indeks untuk list *Synset* dan tabel data untuk data dari *Synset* tersebut. Skema relasi yang akan dibangun ditunjukan pada gambar berikut.



Picture 9 Relation Scheme

2.10.2 Table Structure

The tables contained in the design of the database on the system to be built described as follows.

Table 5 Table Structure t_indeks

Field Name	Data Type	Size	Key	Description
id_indeks	*Integer*	11	PK	*NOT NULL*
lemma	*Varchar*	250		*NOT NULL*
Pos	*Enum*	(‘a’,’n’,’v’,’r’)		*NOT_NULL*
Synset_cnt	*Integer*	3		*NOT NULL*
Ptr_cnt	*Integer*	3		*NOT_NULL*
Ptr	*Varchar*	250		
Sense_cnt	*Integer*	3		*NOT_NULL*
Tagsense_cnt	*Integer*	3		*NOT_NULL*
offset	*Varchar*	250	FK	

Tabel 6 Struktur Tabel t_data

Field Name	Data Type	Size	Kunci	Description
Synset_offset	*Varchar*	8	PK	*NOT NULL*
Lex_filenum	*Int*	3		*NOT NULL*
Ss_type	*Enum*	(‘a’,’n’,’v’,’s’,’r’)		*NOT_NULL*
w_cnt	*Int*	3		*NOT_NULL*
Word	*Text*			
P_cnt	*Int*	3		*NOT_NULL*
Ptr	*Text*			
F_cnt	*Int*	3		*NOT_NULL*
Frames	*Text*			
Definition	*Text*			
Sentence	*Text*			

2.11 Testing Accuracy

In the test the accuracy will be calculated the value of the number of Synset from the result of Unique Strings obtained. The number of Synset available on Wordnet's Princeton is 117659 of all types of word classes. Data extracted from index.noun, index.verb, index.adj, index.adv. Here is the table of extraction results and translation results of the test obtained.

Table 7 Test result

Categories	Preliminary data		Extraction Results		Translated Results	
	Unique Strings	*Synset*	*Unique Strings*	*Synset*	*Unique Strings*	*Synset*
Noun	117953	82192	117953	82192	11944	20992
Verb	11540	13789	11540	13789	5283	10290
Adjective	21499	18185	21499	18185	4755	7366
Adverb	4475	3625	4475	3625	1982	2126
TOTAL	**155467**	**117791**	**155467**	**117791**	**23964**	**40774**

In the test the accuracy will be calculated the value of the amount of Synset from the Wordnet Princeton database. The number of Synset available on Wordnet's Princeton is 117659 of all types of word classes and the total number of Unique Strings 115467. In the test the number of successful Synset obtained using Approach approach with automatic method is 40774 Synset and 23964 Unique Strings

In the first stage of database extraction from PWN obtained 100% accuracy results, all data and its structure successfully obtained. The second phase translates to get the Relation of Meaning and Unique Strings into the target language, to obtain a sysnset development percentage from translation by MRD Oxford Dictionary. Then the calculation performed divides the result of Synset which is successfully translated with the amount of Synset in PWN which is then multiplied by 100%. For details it is as follows.

Number of Princeton Unique Strings Wordnet (US-PWN): 155467

Unique Strings Accuracy

$$= \frac{\text{US-IWN}}{\text{US-PWN}} \times 100\%$$

$$= \frac{153.575}{155.467} \times 100\%$$

$$= 98.78\ \%$$

Research conducted to produce an accuracy of 98.78%

Meanwhile, to get the accuracy of Mean Relation is as follows.

Number of Synnet Princeton Wordnet (S-PWN): 117.791

Synset Accuracy

$$= \frac{\text{S-IWN}}{\text{S-PWN}} \times 100\%$$

$$\text{Antonim} = \frac{47996}{48736} \times 100\% = 98.48\%$$

$$\text{Hipernim / Hyponim} = \frac{59.460}{63660} \times 100\% = 93.40\%$$

$$\text{Meronim / Holonim} = \frac{17.357}{18207} \times 100\% = 95.33\%$$

The results of the study were 98.48% accuracy for antonyms, 93.40% for hyponim / hyponim, and 95.33% for reconstitution / holony.

2.11.1 Analysis of Test Results

Here is an analysis of accuracy test results conducted in the detection of relations between meanings in Wordnet Indonesian.

The resulting data is from the noun, verb, adj, and adverb categories the results obtained for Synset are 117.791 out of 117.791 total and Unique Strings generated 153,575 from 155.467. In the process of translation can be seen in the noun category that has a major impact to the percentage of test results is only obtained about 10% for Unique Strings it is because there is no translation of the translation of the results using the Oxford Dictionary MRD because the word form of noun most is a term noun that is regional in English English. Many Medical terms, Chemistry, and Terms for other specialized dictionaries that can

not be translated and compound words in English are very few of its Indonesian translations.

3. CLOSING

This section describes the conclusions and suggestions of the research conducted.

3.1 Conclusions

The system built in this case is the development of Wordnet Indonesian has been successfully created. And the availability of an open source lexical Database. In the first stage get the result of Unique Strings extraction as much as 155.467 from the total of 155.467 initial data and 117.791 Synset from the total initial data of 117,791. In the second phase of the translation process obtained Unique Strings as much as 153.575 and Synset as much as 117791. Can be seen in the extraction results obtained the results of extraction completely in the first stage, and in the second stage shrinkage results due to the quality of MRD. There are some entries like the word "Knackered" words and a phrase often used by the English people to describe a feeling of fatigue or exhaustion. This word is often used in lieu of the word'exhausted 'which there is no translation and there are still local terms for English that do not exist in the Indonesian language.

3.2 Suggestions

Suggestions for subsequent research are adding meaning elements from KBBI and adding Unique Strings and Synset enhancements by automatic method or Synset development semi-automatically using thesaurus and KBBI. Or use Automatic Translation some MRD then give rank in every translation so get the most appropriate translation.

BIBLIOGRAPHY

[1] D. Novriandi, "Development Indonesian Wordnet," 2016.

[2] E. Pianta, L. Bentivogli, and C. Girardi, "MultiWordNet: developing an aligned multilingual database," *Proc. First Int. Conf. Glob. WordNet*, no. 1996, pp. 293–302, 2002.

[3] J. Ilmiah, I. Komputa, E. Volume, and B. Issn, "SEBAGAI SUMBER DAYA NLP BAHASA INDONESIA Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)."

[4] C. Soler, "Extension of the SpanishWordNet," *Gwc 2004 Second Int. Wordnet Conf. Proc.*, pp. 213–219, 2003.

[5] I. Sommerville, Software Engineering (Rekayasa Perangkat), Erlangga, 2011.

[6] D. A. Chaer, Tata Bahasa Praktis Bahasa Indonesia, Jakarta: Rineka Cipta, 2011.

[7] M. Dr. R Kunjana Rahardi, Bahasa Indonesia Perguruan Tinggi, Erlangga, 2015.

[8] R. beckwith, G. A. Miller and R. tengi, Design and Implementation of the WordNet Lexical Database, 1993.