



PolyWordNet: Analogous to Human Mind for Word Sense Disambiguation

Udaya Raj Dhungana¹ and Subarna Shakya²

^{1,2} Department of Electronics and Computer Engineering, Pulchowk Campus, IoE, Tribhuvan University, Lalitpur, Nepal

Received 29 Nov. 2019, Revised 22 May 2020, Accepted 1 Aug. 2020, Published 1 Sep. 2020

Abstract: PolyWordNet is a new lexical database which deals with the organization of senses of polysemy words. It mimics the way how human mind organizes the senses of polysemy words and their related words to analyze and determine correct meaning of a polysemy word in a context. A related word of a sense of a polysemy word is a word which provides necessary and sufficient context to disambiguate the meaning of the polysemy word. A context with a polysemy word must contain at least one related word that determines the correct sense of the polysemy word. The PolyWordNet utilizes this fact to organize the senses of a polysemy word with their corresponding related words. PolyWordNet is completely different than that of the dictionary and WordNet. The words which spell similar come together in dictionary. The words with similar meaning come together in WordNet. The same words, in WordNet, are connected to the multiple senses of the same polysemy word. This introduces an ambiguity. This ambiguity is resolved in PolyWordNet by linking one related word only with a single sense of the same polysemy word. The PolyWordNet can be used to disambiguate senses of polysemy words more precisely and more efficiently.

Keywords: PolyWordNet, WordNet, Dictionary

1. DICTIONARY AND WORDNET

The dictionary and WordNet are the lexical resources. The information from these resources are used by knowledge-based word sense disambiguation (WSD) methods for sense disambiguation. Lesk Michael in 1986 used the overlap of word definition from the Oxford Advanced Learner's Dictionary of Current English (OALD) to disambiguate the word senses [1]. He used only the definitions of the words that need to be disambiguated from the dictionary.

The lexical information in dictionaries are put together using alphabetical order in which the words that spell alike come together in the list. Keeping the words which spell alike together results in scattering of the words that have similar meanings. For example, the word "put" in the dictionary is scattered with its synonym words "arrange", "place" etc. while the words "pustule" and "put" are together in the dictionary.

To search the words with similar meaning in dictionary is, therefore, tedious and time consuming. To overcome the problems found in lexicographic information, in 1985, a group of psychologist and linguistics was formed in Princeton University to develop a lexical database with the aim to search dictionary conceptually rather than

alphabetically [2]. The resulted product of this research is the WordNet.

The WordNet organizes the words in the lexical database based on their meanings instead of their forms as in dictionaries [3]. It groups the nouns, verbs, adjectives and adverbs together into synonym sets, each expressing a distinct concept [4]. The synonym sets are linked with each other by numerous semantic relations like hyponymy, meronymy, entailment relation etc. [5]. After the development of the WordNet, the lack of information in dictionary is solved since the WordNet contains more information under various semantic relations. WordNet is a popular lexical resource and is massively used for word sense disambiguation [6].

Although the WordNet provides more information under various relations, it still doesn't have any relation that deals with polysemy words. No lexical database, which deals with the relationship of the senses of polysemy words and their related words, is developed till now. Such lexical database is extremely required to resolve the problems in word sense disambiguation. This need motivated us to develop a new lexical database which deals the relationship between senses of polysemy words and their related words.



2. PROBLEMS WITH WSD USING WORDNET

The contextual overlap count knowledge-based WSD approaches use the information from the various relations such as gloss, hypernyms meronyms etc. [7]. Word overlaps between the context and each sense of polysemy word are counted. The sense, which has the maximum overlaps, is taken as the correct sense for the given context.

After the development of WordNet, many WSD methods utilized the information from WordNet for sense disambiguation. Some of these includes: Adapted Lesk algorithm by Banerjee and Pedersen [8], "Weighted Overlapping" disambiguation method by Fragos, Yannis and Christos [9], an unsupervised WSD method by Seo, Hoojung, Hae-Chang, Sung Hyon and Soo-Hong [10], a knowledge-based WSD method by Montoyo, Manuel, Rigau and Armando [11], an overlap-based WSD algorithm for Nepali word sense disambiguation by Roy, Sunita and Bipul [12], word sense disambiguation in queries by Liu, Clement and Weiyi [13], a WSD method for Hindi word sense disambiguation by Sinha, Reddy, Pande, Kashyap and Bhattacharyya [14] and a WSD method for Nepali language by Dhungana and Shakya [15]. These WSD approaches use the information from the various relationships such as synsets, glosses, examples, hypernymy, holonymy, hyponymy, troponymy, meronymy, attribute etc. from the WordNet for sense disambiguation.

The WordNet relations synset, glosses of synset, attribute relation, hypernyms, hyponym, troponym, holonym, meronym, also see, similar to and pertainym, domain relations for nouns, verbs and adjectives have been used to collect the information for sense disambiguation in [7], [8], [9], [10], [11], [12], [13], [15], [16], [17] and [18]. The hyponymy relation does not seem to contribute to the sense disambiguation. The inclusion of definitions from hyponymy relation decreased the accuracy [9], [19].

The higher levels of the WordNet hierarchy are less semantically related than a lower level. The relatives in a synonym class tend to share similar context at higher level hierarchy. For these reasons, use of glosses of relatives of a word in higher level is not appropriate and the definitions in the WordNet still don't contain sufficient information for sense disambiguation. The results from experiments indicate the higher level hypernyms/hyponyms are not useful for all words for sense disambiguation [10], [19], [20]. Very few words are overlapped with context, even the full hypernymy hierarchy from WordNet is used for word sense disambiguation [11]. Information from synonyms, hyponyms, hypernyms, definitions of its synonyms and hyponyms and its domains are not sufficient for sense disambiguation [13], [15]. The WSD method in [21] uses definitions from dictionary but faced the problem of less information in dictionary for word sense disambiguation.

The use of hypernymy from WordNet improves the result of Lesk algorithm. However, when the deeper level hypernyms are used to increase the information for sense

disambiguation, it is found that the accurately disambiguated words are now inaccurately disambiguated. If only first level hypernyms are used, they contain less information for disambiguation. If all level of hypernyms are used, they contain more common information. This common information for each sense does not help to disambiguate rather it introduces a noise information which cause the wrong disambiguation [15].

The noise information is the common information among the senses of polysemy word taken from the WordNet and causes the maximum overlap for a wrong sense of the polysemy word with the given context resulting in wrong disambiguation. More information can be gathered from various relations that are found in WordNet [22], but still it does not provide more distinct information for different sense of the same polysemy words. Even using the deeper level of hypernymy does not provide distinct information that is required to distinguish the different senses of a polysemy word. Here the deeper level hypernyms in WordNet means the hypernyms of a word as we go downwards the hypernym hierarchy in the WordNet.

From these evidences, it can be concluded the information taken from the WordNet are not still sufficient. Sometimes, the information from WordNet create noise information which cause wrong disambiguation [15]. The problems that occur in WSD approaches which use information from WordNet, are described in the following subsections.

A. Insufficient Information in WordNet for Disambiguation

The higher levels of the WordNet hierarchy are less semantically related than a lower level. The relatives in a synonym class tend to share similar context at higher level hierarchy. In some cases, only the gloss of word or hypernym in WordNet is distinct for the senses of the same polysemy word and only such gloss has less information for word sense disambiguation just like in dictionary.

B. Noise Information and Wrong Disambiguation

The higher levels (or from second level) of hypernyms for all senses of the same polysemy word are found to be same in WordNet. This same/common information cannot be used to distinguish the senses of polysemy word. The use of common hypernym induces noise information and this noise information causes wrong sense disambiguation.

C. Disambiguation Depends on Gloss's Words

The overlap count WSD method uses the information from definitions of words in WordNet to count overlaps between context and different senses of polysemy word in the context. The words that are used to define the gloss of a word determines the number of overlaps in these WSD methods. Therefore, the word sense disambiguation using the glosses from WordNet depends on which words are used to define and describe their meaning. This can never be fair for all contexts.



3. SOLUTION APPROACH: POLYWORDNET

This section first presents the way how a human mind analyses the given context to disambiguate the senses of a polysemy word. It then describes how a context (which contains related words) provides a clue to the correct sense of the polysemy word for the given context and how the human mind uses the words in context to disambiguate the sense of polysemy word. Later, it describes how the related words are generated, how these related words are organized and how these words are used for sense disambiguation.

A. Human Mind and Word Sense Disambiguation

Suppose a context "She is eating bass". Here, the word *bass* is polysemy word and its meaning needs to be disambiguated. When a human mind reads this context, it is so intelligent that the mind finds the relation between *eating* and *bass*. It analyses these two words and concludes that what a human can eat is a *bass* fish. The word *eating* is a sufficient evidence for human mind to conclude the *bass* is a fish. These two words *eating* and *bass* are so connected and stored in human mind that they have strong relationship to disambiguate the meaning of the polysemy word *bass* to be a *bass* fish. Suppose the same context with one more word- "She is eating bass with spoon". In the context, another word *spoon* has the relation with *eating* and *eating* has the relation with *bass*. The *spoon* is used to eat something and *bass* fish can be eaten. These two words *spoon* and *eating* are so connected and stored in mind that the mind can conclude "spoon is used to eat something eatable". When the human mind reads this context, it finds the relation of *spoon* with *eating* and concludes so fast that what can be eaten with spoon is a *bass* fish. Here, the word *spoon* is supporting for the human mind to conclude the eaten *bass* in the given context is a fish. This is a way how a human mind analyzes the context and understands the correct meaning of a polysemy word. We call these words *eating* and *spoon* as related words for the sense "a fish" of polysemy word *bass*.

Suppose a context "John likes bass". There is no any related word which is sufficient to disambiguate the correct sense of the bass. Even human cannot disambiguate the meaning in this context. John may like a bass fish or bass music. At least a sufficient related word must be provided even for human to disambiguate the meaning of a polysemy word.

In human mind, the related words and the senses of polysemy words in a context are so connected and stored that when human reads the context containing a polysemy word, the human mind finds the related words, analyses and connects these related words with the correct sense of the polysemy word in the given context. Motivated from this organization of senses of polysemy words and their corresponding related words in human mind, a new lexical database called PolyWordNet is developed to organize the senses of polysemy words and their corresponding related words.

B. Polysemy Words and Related Words

A polysemy word has multiple meanings according to the contexts where it is used. The context determines the exact meaning of the polysemy word. Therefore, without a sufficient context, even a human cannot determine the correct meaning of the polysemy word. Therefore, a given context must be sufficient to disambiguate sense of polysemy word. The words which determine the correct sense of a polysemy word in the given context are called related words of that sense.

Suppose a context "Maria is writing a poem with my pen". In this context, the word *writing* is a related word which determine the sense of polysemy word *pen* as a writing implement. In addition, another word *poem* is also a related word to that sense since what can be used to write a poem is a writing implement (pen). These two related words provide sufficient context to human mind to determine the correct sense of the *pen*.

Based on this fact, if the senses of polysemy words are semantically connected with their corresponding related words, the resulted lexical database can be used for word sense disambiguation to get exceptionally higher accuracy. Such lexical database can be used just like a human mind for word sense disambiguation.

In dictionary, the words which spell similar comes together. This organization doesn't link up the senses of polysemy words and their corresponding related words. Another lexical database WordNet organizes the words based on synonym set. It brings the words with similar meaning together and provides more information about a word. However, it still does not care about the relation between the senses of polysemy words and their corresponding related words.

This research work strongly believes that a given context always contains at least a related word for a polysemy word. Therefore, if each sense of polysemy words is connected with their corresponding related words, these relations can be used to completely resolve the problem of word sense disambiguation in simple sentences containing a single polysemy word.

Polysemy word: <i>Pen</i>		Sense: <i>Writing Implement</i>
SN	RULES	RELATED WORDS
1	All possible attributes.	black, red, green etc.
2	All possible functions.	write
3	All entities it contains or all its constituents parts.	cap, nib, ink etc.
4	All entities with which its functions are related.	poem, song, homework etc.
5	All entities with which it is used for.	-
6	All entities that describe it or its function or way of doing its function.	-
7	All entities along with it occurs.	book, copy, note, pencil, bag etc.

Figure 1. Generating related words for the sense "writing implement" of polysemy word *pen*.



C. Generation of Related Words

The related words are the main key for the word sense disambiguation using PolyWordNet. Finding a set of good related words is a difficult task. A good related word must possess two essential features. Firstly, it must not lead to create another ambiguity during sense disambiguation process. Secondly, a related word(s) must provide context to the sense. To determine the related words for a sense of a polysemy word, take the sense as an entity and find out the following information (if applicable) for that sense:

- All possible attributes.
- All possible functions.
- All entities it contains or all its constituent parts.
- All entities with which its functions are related.
- All entities with which it is used for.
- All entities that describe it or its function or way of doing its function.
- All entities along with it occurs.

For example, Fig. 1 shows how to generate the related words for the sense "writing implement" of polysemy word pen. From the Fig. 1, the related words for the sense "writing implement" of polysemy word *pen* are black, red, green, write, cap, nib, ink, poem, song, homework, book, copy, note, pencil and bag. All the generated sets of related words of each sense of a polysemy word are checked to find whether they contain common words. If they contain common words for all or some senses of a polysemy word, those words are removed.

D. PolyWordNet: Net of Senses of Polysemy Words and Related Words

Analogous to the human mind, the PolyWordNet organizes the senses of polysemy words and their corresponding related words in such a way that each sense of a polysemy word is inter-connected with its all possible related words. In this organization of words, the sense of a polysemy word and its related words come together and form a cluster.

Organization of Words

The related words are inter-linked with their corresponding senses of polysemy words. The PolyWordNet organizes multiple senses of a polysemy word in such a way that each sense of the polysemy word is linked with its related words by dividing these related words into verbs, nouns, adverbs and adjectives.

In PolyWordNet, each related word is linked only with a sense of a polysemy word. If a word is equally semantically related with more than one sense of the same polysemy words, it is just ignored and is not included in the related words of either sense. Let us consider the three senses of word "pen":- pen 1- "a writing implement with a point from which ink flows", pen 2- "an enclosure for containing livestock" and pen 3- "a portable enclosure in which babies may be left to play". Also let the three senses "pen 1", "pen 2" and "pen 3" have the three sets of related words as {copy, book, poem, write}, {rabbit, dog} and {doll, baby} respectively. Then, these related words are linked with their respective senses as shown in Fig. 2.

Word Sense Disambiguation

The developed new WSD algorithm, which uses relations from PolyWordNet for sense disambiguation, does not count the number of overlapped words between the context and sense information. Instead, this algorithm searches the paths or links of context words with a sense of a polysemy word that needs to be disambiguated in a given context. If the paths thus obtained connect the context words only with one sense of the polysemy word, the algorithm outputs the linked sense as the correct sense of the polysemy word. If there are paths that link more than one senses, then the algorithm counts the number of paths or links for each linked sense.

The sense for which the number of connection paths is maximum is selected as a correct sense. If no connection is found, the algorithm displays an information indicating the failure in disambiguation. Fig. 4 shows the new algorithm which uses PolyWordNet for sense disambiguation.

PolyWordNet Database

The main rationale behind the development of PolyWordNet is to deal with and organize the senses of a polysemy word and their related words. To build the PolyWordNet database, both the senses of polysemy words and related words are stored in a main table called "word_info" in database and other tables are used to store the relations between words (i.e. links between the related words and corresponding senses of polysemy words). These tables that are used to store the relation links among words includes "noun", "verb", "adjective" and "adverb". Different tables are used to store the relation links for nouns, verbs, adjectives and adverbs. The Fig. 3 shows the database diagram of lexical database PolyWordNet.

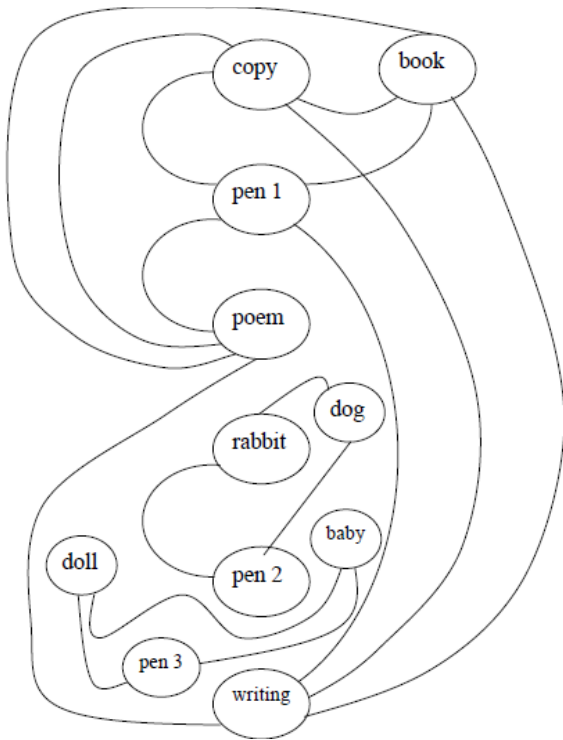


Figure 2. Organization of senses of polysemy words and their corresponding related words in PolyWordNet.

E. Data Generation and Test Data

The data required in this research mainly includes the data to build the PolyWordNet and Test Data. To build the PolyWordNet, it requires senses of polysemy words and their corresponding related words. The first task was to collect polysemy words and their multiple senses. To collect polysemy words, the popular WordNet and different online dictionaries in websites are used. For this, undergraduate students who are studying at Computer Science course in different Universities- Tribhuvan University, Nepal, Pokhara University, Nepal and Darmstadt University, Germany are chosen.

They collected the polysemy words along with different senses mainly from WordNet and online dictionaries. To prepare test data, those respondents are request to build the sentences for each sense of polysemy words or collect from the web resources. They also asked to collect the related words for each sense of polysemy words using the sentences available from the web resources.

They collected altogether 3541 different words to build the PolyWordNet. They also collected 2905 sentences to build Test Data. Out of 3541 words, 1748 are polysemy words and 1793 are single sense words. Any word can be a related word of a single sense of the same polysemy word. However, the same word can be the related word of single sense of many polysemy words. A sense of a polysemy word can be a related word of another polysemy word.

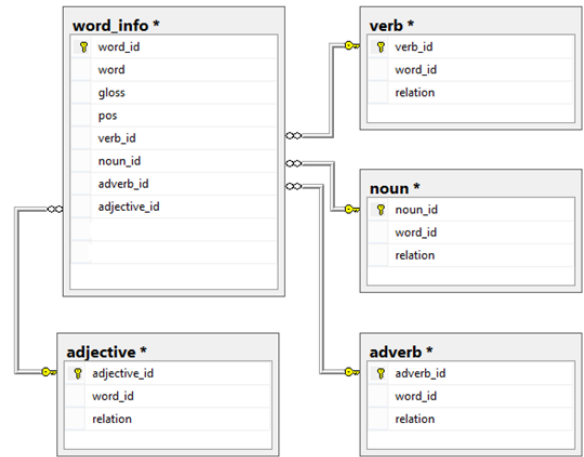


Figure 3. PolyWordNet Database.

The same words are used in sample WordNet to keep the constant information in both lexical databases. In addition, in WordNet, all the hypernyms, hyponyms and meronym are used directly from the WordNet during sense disambiguation. In this sense, the WordNet contains more information including the relations like, synset, hypernyms hyponyms and meronyms. However, the total number of words in WordNet are made same as in PolyWordNet.

Altogether 42 different experiments are run using the relations from WordNet and PolyWordNet containing altogether 3541 different words and these experiments are tested using 2905 different simple sentences generated in this research and 100 sentences from *news* category of Brown corpus. Here simple sentences means those sentences which are not compound sentences and are not ambiguous to human.

4. MATHEMATICAL MODEL

In PolyWordNet, the senses of a polysemy word and their corresponding related words are put together and they form a group. The PolyWordNet and the new WSD algorithm can be mathematically formulated as:

A. PolyWordNet

Let us consider *pw* be a polysemy word. Assume *pw* has *n* different senses $S_1, S_2, S_3... S_n$. Each S_i (where $i = 1, 2, 3... n$) has some related words. Suppose $RW (S_i)_k$ is a collection of related words for sense S_i of polysemy word *pw*. Then, PolyWordNet *P* is a collection of relations (say *Rel*) between S_i and $RW (S_i)_k$ of all polysemy words *pw* and is defined as

$$RW (S_i)_k = \left\{ W \mid \begin{array}{l} W \text{ is a word} \\ \text{related with } S_i \end{array} \right\} \tag{1}$$

$$Rel = \left\{ \begin{array}{l} (S_i, RW (S_i)_k) \\ \in P \end{array} \mid \begin{array}{l} \text{each } W \in \\ RW (S_i)_k \text{ is} \\ \text{connected} \\ \text{with } S_i \end{array} \right\} \tag{2}$$



$$P = \{x|x \text{ is a relation Rel}\} \quad (3)$$

B. Word Sense Disambiguation

Let us consider, in a given sentence, pw is a polysemy word and has n different senses $S_1, S_2, S_3 \dots S_n$. The correct sense of the polysemy word pw needs to be determined using the context of a given sentence. Other remaining words in the sentence provide the context for the polysemy word.

```

ALGORITHM: Word Sense Disambiguation using PolyWordNet

INPUT: A given context and a target word (polysemy
word whose meaning is going to be
disambiguated in the given context).

OUTPUT: A sense of the target word (polysemy word).

Prepare context words by removing articles and prepositions
Store the target word in a variable
For each word in the context words
  start traversing with a context word
  If it reaches one of the sense of target word
    store the path and stop the traversing
  Else
    traverse through all possible paths found in PolyWordNet
  End if
End For

If number of connections is zero
  Output message "Disambiguation failed"
  Exit
Else
  For each traversed path
    If there are connections with only one sense of target word
      output the connected sense as correct sense
      exit loop
    Else
      Count the number of connections for each connected senses
      If number of max. connections for two senses are equal
        Output the first among the two senses as correct sense
        exit loop
      Else
        Output the sense with max connections as correct sense
        exit loop
      End if
    End if
  End For
End if

```

Figure 4. Word Sense Disambiguation using PolyWordNet.

Suppose, there are k different context words CW_j where $j = 1, 2, 3 \dots k$ in the sentence. The correct sense S_i is determined if there exists a function $f(CW_j, S_i)$ in PolyWordNet P such that any one context word CW_j is in $RW(S_i)_k$ and there is a path from CW_j to S_i . The function $f(CW_j, S_i)$ is defined as

$$f(CW_j, S_i) = \begin{cases} 1 & \text{if } CW_j \in RW(S_i)_k \text{ and path } (CW_j \text{ to } S_i) \\ 0 & \text{Otherwise} \end{cases} \quad (4)$$

5. EXPERIMENTAL DESIGN

The hypothesis of this research is that if there is a polysemy word in a context, the context also contains the related word. This related word is sufficient evidence to prove the correct sense of the polysemy word in the context.

This relationship between senses of polysemy word and their corresponding related words can be used to organize the sense of polysemy words. Thus formed lexical database that organizes the senses of polysemy words and their

related words can be used to increase the accuracy of word sense disambiguation.

All together six series of experiments are set up and executed to test the formulated hypothesis. The number of data in each series is increased to observe whether the increase in data increases the accuracy or not for overlap count knowledge based approaches that uses the information from WordNet for word sense disambiguation.

Each series has 4 different experimental set up. The first 3 experiment has two runs- run A (or run 1) and run B (or run 2). Hereafter, the run A and run 1- the both mean the same run. Similarly, run B and run 2 mean the same run. These terms can be interchangeably used to mean the same experiment run. The difference between run A and run B is that the sense bag in the run B of each experiment contains the hypernym of every words of sense bag used in the run A of each experiment. Thus, the run B of each experiment always contains more information than the run A. These 3 experimental settings (each containing 2 runs, thus altogether 6 different experiments) uses the information from WordNet for word sense disambiguation.

These experimental settings are named as 1) Exp 1 Run A, 2) Exp 1 Run B, 3) Exp 2 Run A, 4) Exp 2 Run B, 5) Exp 3 Run A and 6) Exp 3 Run B. In contrast to these experiments, the experiment 4, uses the relations from PolyWordNet for word sense disambiguation. This seventh experiment is named as Exp 4. Thus, every series of experiment contains 7 different experimental settings and thus 6 series of experiments contain altogether 42 different experimental settings.

The intent of these experiments is to compare and observe which lexical database WordNet or PolyWordNet is better for word sense disambiguation. A sample information was taken. From this information both sample WordNet and PolyWordNet were built. Same information were used to build both lexical databases in order to keep the amount of information constant throughout the experiments.

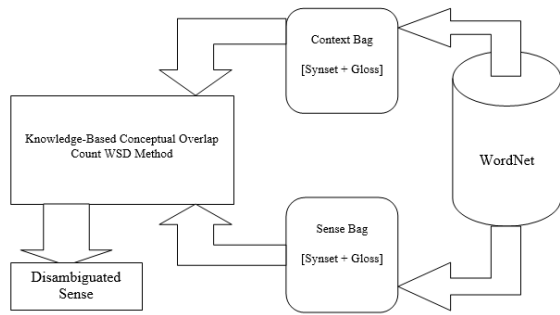
The details of these four experiments are described in the following subsections.

A. Experiment 1 - Exp 1 Run A and Exp 1 Run B

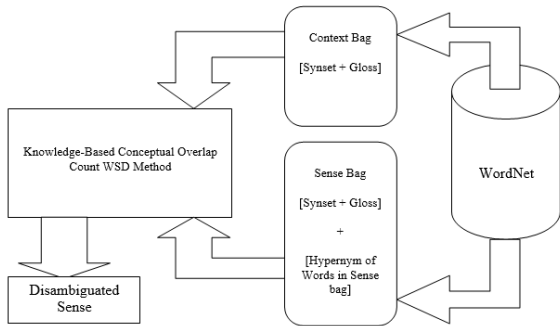
Rationale: The rationale behind carrying out this experiment is that the information taken only from the gloss of words from the WordNet is insufficient for sense disambiguation.

Intent: This experiment is designed with an intent to represent the knowledge-based contextual overlap count WSD method that uses the information only from synset and gloss in WordNet. Each experiment is run twice.

In first run- **Exp 1 Run A**, only the synset and gloss of words in sense bag and context bag are used. In the second run- **Exp 1 Run B**, the hypernym of words in each sense bags for every experiment are included to observe the effect of increasing the information from hypernyms of words in sense bag.



(a) Exp 1 Run A



(b) Exp 1 Run B

Figure 5. Experiment 1 Setting (a) Exp 1 Run A and (b) Exp 1 Run B.

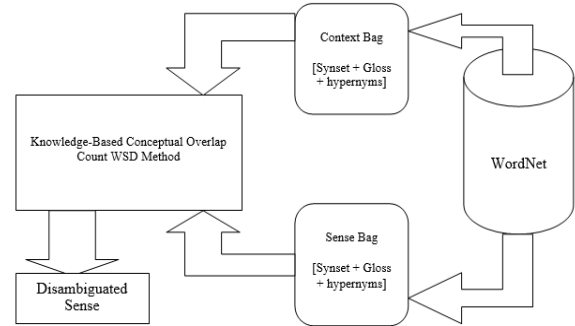
Experimental Setting: For this experiment, simplified Lesk algorithm is used. This algorithm used the information from only from *synset and gloss* in the WordNet to form the sense and context bags.

B. Experiment 2 – Exp 2 Run A and Exp 2 Run B

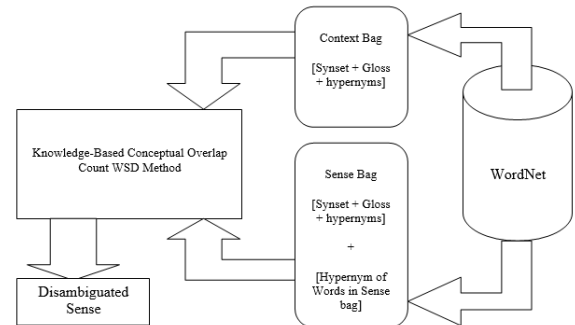
Rationale: The rationale behind carrying out this experiment is that if the information in the sense bag and context bag is increased, this will increase in the relatedness of correct sense with the context.

Intent: This experiment is designed with an intent to represent the knowledge-based contextual overlap count WSD method that uses the information from synset, gloss and hypernyms in WordNet.

In first run- **Exp 2 Run A**, the information from synset, gloss and hypernyms in WordNet are used. In the second run- **Exp 2 Run B**, the same information plus hypernym of words in each sense bags are included.



(a) Exp 2 Run A



(b) Exp 2 Run B

Figure 6. Experiment 2 Setting (a) Exp 2 Run A and (b) Exp 2 Run B.

Experimental Setting: For this experiment, the simplified Lesk algorithm is used. This algorithm used the information only from *synset, gloss and hypernyms* in the WordNet to form the sense and context bags.

C. Experiment 3 – Exp 3 Run A and Exp 3 Run B

Rationale: The rationale behind carrying out this experiment is that if the information from hyponyms and meronyms in the sense bag and context bag is increased, this will further increase in the relatedness of correct sense with the context.

Intent: This experiment is designed with an intent to represent the knowledge-based contextual overlap count WSD method that uses the information from synset, gloss, hypernyms, hyponyms and meronyms in WordNet. In first run- **Exp 3 Run A**, the information from synset, gloss, hypernyms, hyponyms and meronyms in WordNet are used.

In the second run- **Exp 3 Run B**, the same information plus hypernym of words in each sense bags are included.

Experimental Setting: For this experiment, simplified Lesk algorithm is used. This algorithm used the information from only from *synset, gloss, hypernyms, hyponyms and meronyms* in the WordNet to form the sense and context bags.

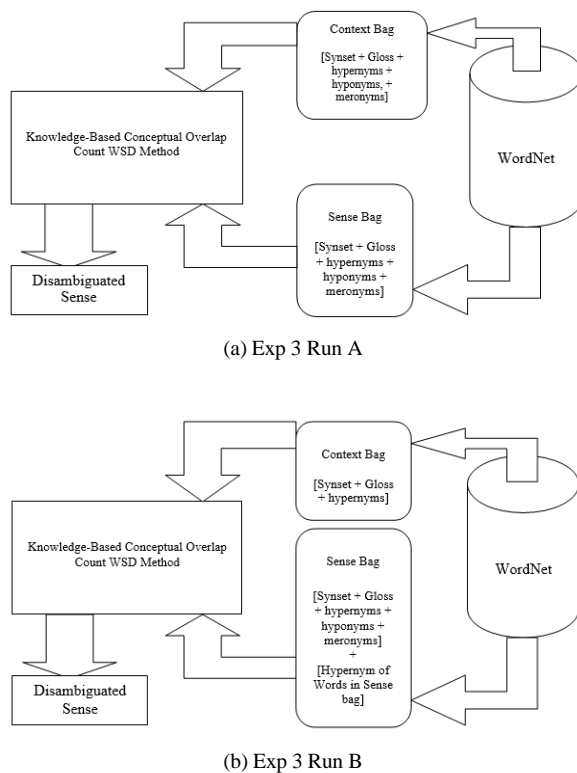


Figure 7. Experiment 3 Setting (a) Exp 3 Run A and (b) Exp 3 Run B.

D. Experiment 4 – Exp 4

Rationale: If the senses of polysemy word and the contextual related words are inter-linked, it removes the unwanted noise information that cause the wrong disambiguation of sense. The removal of this noise information increases the accuracy of the WSD approaches.

Intent: The intent of this experiment is to show if the senses of a polysemy word and the related words with each sense of the polysemy word are inter-linked to each other, it resolves the ambiguity by eliminating the unwanted noise information and to check whether the new WSD algorithm (one of the deliverable of this research) using the direct inter-linked relations from the lexical database PolyWordNet (another deliverable of this research) obtains higher accuracy for sense disambiguation than that of the algorithm using the contextual overlap count WSD methods using the WordNet.

The second intent of this experiment is to check whether the related words collected from the same information available from the relations in WordNet (that is used in experiment 1, 2 and 3 by knowledge-based contextual overlap count WSD methods for sense disambiguation) when linked with the corresponding senses of polysemy word in PolyWordNet will result in higher accuracy than the accuracy obtained in experiment 1.

If the accuracy of this experiment is found to be higher than that of the accuracy in experiment 1, it proves that the same information when arranged in PolyWordNet yields higher accuracy than that of the use of WordNet.

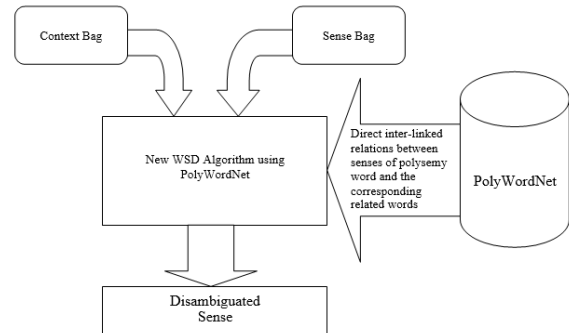


Figure 8. Experiment 4 Setting (Exp 4).

Experimental Setting: In this experimental setting, the WordNet is replaced by our new lexical database PolyWordNet. The simplified Lesk algorithm is also replaced with new WSD algorithm which uses the direct inter-linked relations from PolyWordNet for sense disambiguation.

6. BUILDING SAMPLE WORDNET AND NEW LEXICAL DATABASE- POLYWORDNET

The PolyWordNet, which is developed currently, has few words. It contains 3541 words including nouns, verbs, adverbs and adjectives. It will be unfair to compare the usefulness of WordNet which contains a huge amount of information. Therefore, a sample WordNet is built so that the both lexical databases WordNet and the PolyWordNet have the same amount of information for word sense disambiguation.

The intent of these experiments is to observe the results by comparing the WSD algorithms that use WordNet and PolyWordNet so that we can conclude which lexical database has a better word organization for word sense disambiguation.

To keep the same environment and conditions, a sample WordNet is built in the same principle as it was developed at Princeton University. A sample information was taken and the both sample WordNet and PolyWordNet were built. Thus, these two lexical databases WordNet and PolyWordNet contain the same information but they have different word organization.

The amount of information in WordNet and the PolyWordNet is the extraneous variable. The extraneous variable is kept constant throughout the experiments by building the both lexical resources the WordNet and PolyWordNet from the same amount of information.

7. RESULT ANALYSIS

To test the hypothesis, 6 series of experiments are set up and run. Each series contains 7 different experimental



setups named as Exp 1 Run A, Exp 1 Run B, Exp 2 Run A, Exp 2 Run B, Exp 3 Run A and Exp 3 Run B and Exp 4.

In each of these 7 experiments, the information in context bag and sense bag are increased from experiment one to four and from run A to B. In addition, the number of words in lexical databases WordNet and PolyWordNet is increased to observe the effects of increase in number of words in those database.

A. Series A Experiments and observed results

The **Series A Experiments** are tested only with 280 words stored in PolyWordNet and Sample WordNet. The statistics of the words used in Series A Experiments are shown in Table I.

TABLE I. WORDS STATISTICS USED IN SERIES A EXPERIMENTS

Total Words	Polysemy Words	Single Sense Words	Nouns	Verbs	Adverbs	Adjectives
280	56	224	213	56	1	10

Fig. 9 shows the accuracies obtained in all Series A experiments. In experiment 1, 77 out of 180 test sentences are correctly disambiguated in the first run in which only the synset and gloss of the words are used to form the context and sense bags. For this case, the accuracy is found to be 42.78%. For the second run, the information in each sense bag is increased by including the hypernyms of words in the gloss of each sense. When the second run of the experiment 1 is run, it is found that 78 out of 180 test sentences are correctly disambiguated. This time, the correctly disambiguated test sentences is increased only by one giving the accuracy of 43.33%. This is not obviously a significant increase in the accuracy to say the accuracy is increased when the information is increased by including the hypernym of the words in the gloss of senses in the sense bags. The 13 test sentences out of 77 which were correctly disambiguated in first run A, are incorrectly disambiguated in second run B when the information in the sense bag is increased. Similarly, 14 test sentences which were incorrectly disambiguated in first run are correctly disambiguated in the second run. This is shown in Fig. 9.

In this case, there is no significant increase in the correct disambiguation with increase in the information in sense bag. The most important point is that correctly disambiguated sentences in first run are incorrectly disambiguated in the second run when the information is increased. This indicates that only increasing the information and using the hug information from WordNet does not ensure the correct disambiguation of sense.

In this case, there is no significant increase in the correct disambiguation with increase in the information in sense bag. The most important point is that correctly

disambiguated sentences in first run are incorrectly disambiguated in the second run when the information is increased. This indicates that only increasing the information and using the hug information from WordNet does not ensure the correct disambiguation of sense.

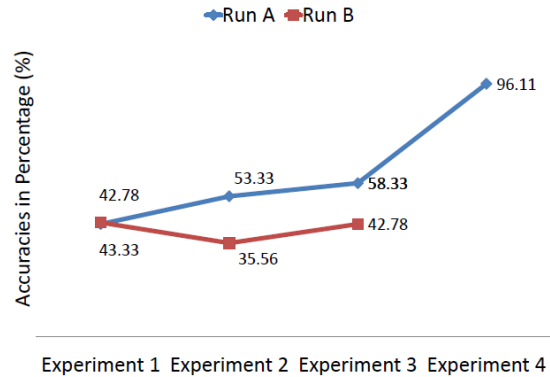


Figure 9. Accuracies obtained in Experiments.

The result of first run A of Experiment 2 shows that 96 test sentences out of 180 are correctly disambiguated while the 64 test sentences are correctly disambiguated in second run B of the experiment 2. The accuracy is found to be 53.33% in first run A and 35.56% in second run B. Comparing the first run of experiment 2 with the first run of experiment 1, it is observed that the accuracy is increased by 10.55% in experiment 2 when the information from hypernyms are included to form the context and sense bag. This indicates that the information from hypernym are useful for the sense disambiguation. However, when the information from hypernyms of the words in each sense bag is further increased in second run, the accuracy is found to be decreased to 35.56% from 53.33% by 17.77% which was much unexpected. If the accuracy of second run of experiment 2 is compared with the second run of experiment 1, it is found that the accuracy is decreased from 43.33% to 35.56% by 7.77% when the information is increased from hypernyms in context and sense bags.

The most important point noted here is that increasing the information in context and sense bag from WordNet does not always increase the relatedness between correct sense of a polysemy word and the context. It is found that the information from WordNet introduces the noise information which causes wrong disambiguation of sense of polysemy word. Here, the noise information means the information that is included from relations of WordNet which when included in the context or sense bag increases more general information and causes a highest overlap between wrong sense of polysemy word and the context resulting in the wrong disambiguation of sense.

Let us see the Fig. 10. In case of Experiment 2, the number of correctly disambiguated test sentences in run A but incorrectly disambiguated in run B (i.e. C→W) is 57 which is very high as compared with the runs of experiment 1. This means the increase in the information in sense bag



induces more noise information which leads to wrong disambiguation of sense. The C→W is even higher in experiment 3. In experiment 3, the information is further increased in context and sense bags by including synset, hypernyms, hyponyms and meronyms. At this time, the 105 test sentences out of 180 are correctly disambiguated giving the accuracy of 58.33% in first run while only the 77 test sentences are correctly disambiguated in second run. The Fig. 10 shows the effects of increase in information in context and sense bags in these three experiments.

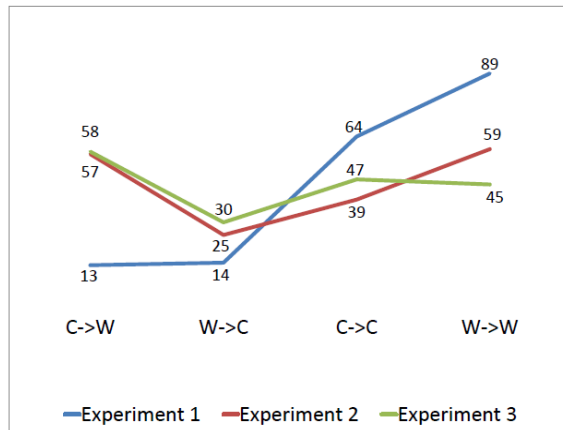


Figure 10. Effect of increase in information in context and sense bags in the first three experiments.

Comparing with the first run of experiment 3 with the first run of experiment 2, it is found that the accuracy is increased only by 5% in first run of experiment 3. The information in second run is more than the first run for every experiment. However, the accuracy of second run of experiment 3 is greatly decreased (i.e. 42.78%) than that of first run of experiment 3. With increase in information in sense bag, the accuracy is decreased to 42.78% from 58.33% by 15.55%. From these observations, it is clear that only increasing the information in context and sense bags from WordNet does not increase the accuracy. Sometimes, this increase in information decreases the accuracy. Another most important point is that the effect of increase in the information from WordNet for sense disambiguation is not consistent.

The WordNet contains the very useful information required for natural language processing tasks. However, the general information, when used, causes the wrong word sense disambiguation. From the results, it is observed that the information only from the definitions of the words contain less information as in first run of Experiment 1. When more information are used from hypernyms as in first run of experiment 2, there will be more overlaps for correct sense. Therefore, the accuracy is increased to 53.33% by 10.55% from 42.78%. However, when there is an increase in the amount of information from hypernyms of words in each sense to make more information in each sense bag as in second run of experiment 2, due to the entry of more common information for all senses, correctly

disambiguated words are also now incorrectly disambiguated and the accuracy is decreased to 35.33% by 17.77%.

This evidence can also be compared with the work by Fragos, Yannis and Christos [9]. They also tested their system including the hyponymy relation but the experiment showed that there was no improvement in accuracy using the hyponymy. This is also supported by the work of Dhungana and Shakya [15]. They found in their experiment that when deeper levels of the hypernyms are used, the correctly disambiguated polysemy words are also incorrectly disambiguated. In such case, there is no meaning of using the hypernymy of the two different senses of the polysemy word to disambiguate their meanings. From the results of the first three experiments, it is proved that the inclusion of hypernyms or many levels of hypernyms do not provide such distinct information which is necessary and sufficient for sense disambiguation.

The noise information in context or sense bag is produced in such a way that more common information are inserted into context bag and into any of the incorrect sense of a polysemy word making the more overlaps between the context and the incorrect sense of the polysemy word. This induced noise in the context and incorrect senses of polysemy word causes the incorrect disambiguation of the sense.

The experimental settings of the first three experiments are similar since all these three uses the WordNet and simplified Lesk algorithms for sense disambiguation. The experimental setting of experiment 4 is completely different. The new lexical database PolyWordNet is used in experiment 4. In addition, new WSD algorithm is used in experiment 4 for sense disambiguation. The 173 test sentences out of 180 are correctly disambiguated giving the high accuracy of 96.11% which is significantly higher than the accuracies found in the first three experiments. In this experiment only the 7 test sentences are found to be incorrectly disambiguated. The reason behind the higher accuracy obtained in experiment 4 is as follows: the new lexical database PolyWordNet organizes multiple senses of a polysemy word in such a way that each sense of the polysemy word is linked with its related words. In WordNet, the words that are used as context words, are connected to the multiple senses of a polysemy word. In this condition, such words cannot be used to disambiguate the meaning of the senses of polysemy word since they are related with more than one sense of the same polysemy word. This condition introduces the ambiguity in ambiguity. This is resolved in PolyWordNet by linking one related word only with a single sense of a polysemy word.

B. Series B, C, D and E Experiments and Obtained results

The main intend to run these **Series B to Series E** Experiments is to observe the effect on results of the 7 experiments (within each series) on increasing the number of data in lexical databases- PolyWordNet and sample



WordNet. The **Series B Experiments** are run on **290** words, **Series C Experiments** on **1477** words, **Series D Experiments** on **2501** words and **Series E Experiments** on **3541** words. This is shown in Table II.

In addition to the increase in data, in each series of experiments, the number of Test Sentences (TS) in Test Data is also increased except for Series B Experiments. The **Series B Experiments** are tested with **180 TS**, **Series C Experiments** with **930 TS**, **Series D Experiments** with **1930 TS** and **Series E Experiments** with **2905 TS**.

The intend to increase the number of TS in each series of experiments is to observe the effects on results when the experiments are tested with the new set of TS. The total number of words used in the Series E Experiments are 3,541. The detail word statistics of the Series E Experiments is shown in Table II.

TABLE II. WORDS STATISTICS USED IN SERIES B, C, D AND E EXPERIMENTS

Total Words	Polysemy Words	Single Sense Words	Nouns	Verbs	Adverbs	Adjectives
3541	1748	1793	2264	859	51	367

The total occurrences of Polysemy words are 1,748 while the number of single sense words are is 1,793.

The accuracies of the experiments Exp 1 Run A, Exp 1 Run B, Exp 2 Run A, Exp 2 Run B, Exp 3 Run A and Exp 3 Run B are not consistent with the increase in number of data (words) in sample WordNet (see Table III). The accuracies are found to have falls and rises with the increase of data. The highest accuracy obtained is 60% in experiment **Exp 3 Run B** in Series B Experiments. The experiment **Exp 3 Run B** in Series E has uses the maximum number of data for sense disambiguation. However, its accuracy is 57.04% which is less than the accuracy of **Exp 3 Run B** in Series B. This evidence clearly indicates that the increase in data in WordNet doesn't ensure the increase in accuracy. Rather it indicates that the increase in data sometimes cause to decrease the accuracy.

TABLE III. OBSERVED ACCURACIES OF SERIES B, C, D AND E EXPERIMENTS

Exp Series	No of Words	No of TS	Accuracies Obtained (in Percentage)							Exp 4
			Using WordNet						Using PolyWordNet	
			Exp 1 Run A	Exp 1 Run B	Exp 2 Run A	Exp 2 Run B	Exp 3 Run A	Exp 3 Run B		
Series B	290	180	46.11	55.55	48.88	52.22	51.11	60	98.33	
Series C	1477	930	41.61	52.68	53.01	49.46	54.4	53.11	98.7	
Series D	2501	1930	45.39	57.20	51.71	53.16	52.80	56.89	99.17	
Series E	3541	2905	44.99	57.83	51.43	52.36	51.33	57.04	99.24	

The result of **Exp 4** shows the rise in accuracy with the increase in data throughout each successive series. The lowest accuracy is obtained in Series A. It goes on increasing and the highest accuracy is found in Series E. There is no fall in accuracy with the increase in data in PolyWordNet. The Fig. 15 shows the accuracies of Exp 4 obtained in Series A to E. This shows a slight rise in accuracy which indicates that increase in number of related words increases the accuracy.

The Fig. 11 shows the accuracies obtained in all experiments from Series B to E experiments. It clearly shows that the accuracy of all experiments that uses WordNet tend to form a constant line with maximum accuracy 60%. This indicates that the even increasing the data in WordNet for disambiguation there is no sudden change in accuracy. This means the amount of data represents the accuracy in whole population as well. The Fig. 12 shows the accuracy obtains by only 30 different experiments which are using WordNet. The minimum accuracy obtained among these 30 experiments is 35.56% in experiment **Exp 2 Run B** in Series A Experiments. The maximum accuracy obtained among these 30 experiments is 60% in experiment **Exp 3 Run B** in Series B Experiments. The average accuracy of all these 30 experiments is 50.88%. This also shows that the accuracies of Exp 4 in successive Series A to E are relatively high.

The Fig. 13 shows the falls and rise in accuracy of Series D experiments which uses the WordNet. The Fig. 14 shows the falls and rises in accuracies of Exp 2 Run 2 (Exp 2 Run B) in each successive series. This indicates that with the increase in information from WordNet does not always guarantee to increase the accuracy.

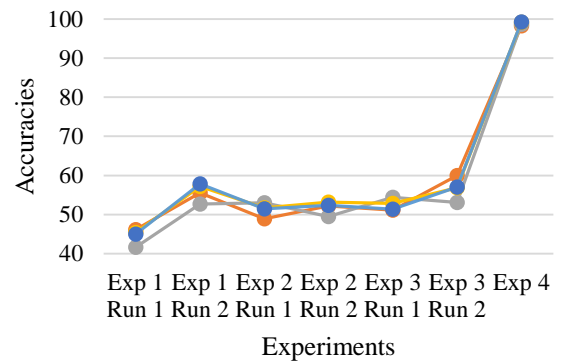


Figure 11. Accuracies observed in Series B, C, D and E experiments.

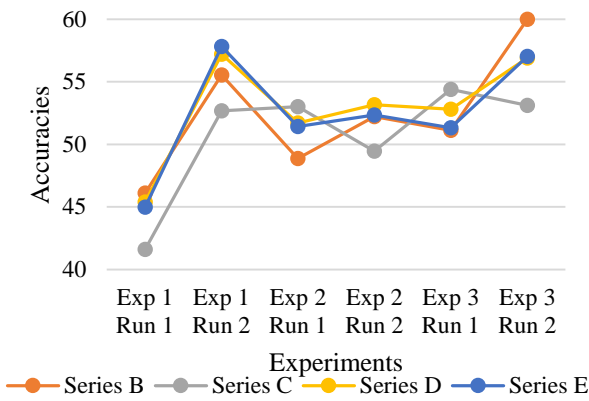


Figure 12. Accuracies observed in Series B, C, D and E experiments (Series-wise).

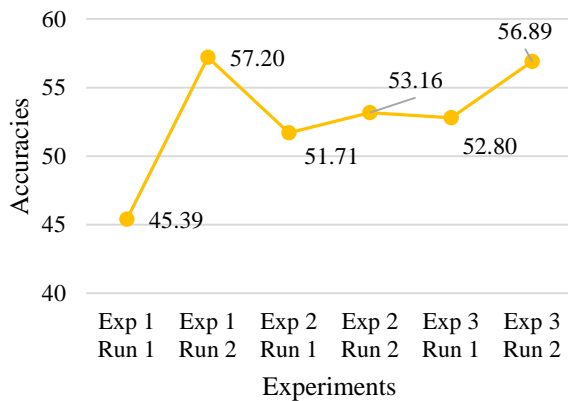


Figure 13. Accuracies observed in Series D experiments that use information from WordNet.

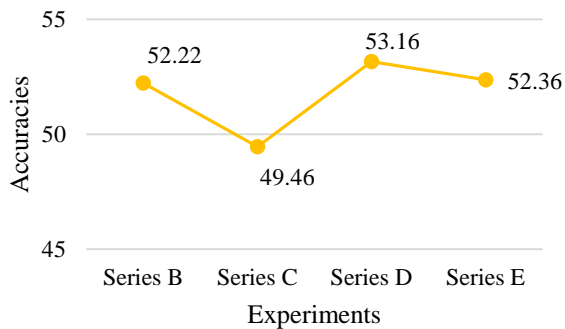


Figure 14. Falls and Rise in accuracies of Exp 2 Run B (Exp 2 Run 2) that uses information from WordNet.

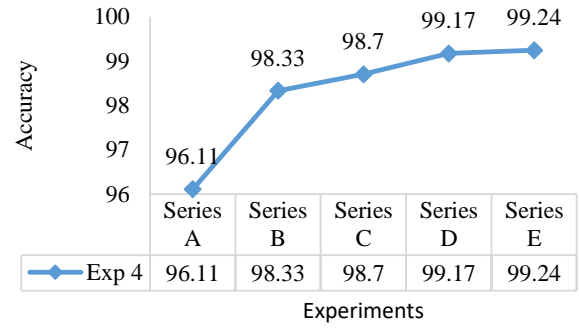


Figure 15. Observed accuracies of Exp 4 that uses relations from PolyWordNet.

The remaining 5 experiments (Exp 4 in each series) uses PolyWordNet for word sense disambiguation. The minimum accuracy of experiments using PolyWordNet is 96.11%. The maximum accuracy is 99.24% and the average accuracy is 98.31%.

The results of experiments indicates that for a particular experiment (say Exp 3 Run A), there is no significant differences in accuracy as the data are increased throughout the successive series. There is much less difference in accuracy in successive series. These facts of results indicates that the accuracy of WSD methods using either PolyWordNet or WordNet remains almost same even if the data are increased in these lexical database. Finally, the results of these experiments proves the PolyWordNet gives better accuracy than the WordNet for word sense disambiguation.

C. Series F Experiments and Observed Results

The Series F experiments is tested on 3541 words. The A to E series experiments are tested by 2905 test sentences generated in this research from web. The Series F experiments are tested by 100 test sentences randomly taken from *news* category of Brown corpus. The 24 sentences out of 100 are simple sentences and remaining 76 sentences are compound and highly ambiguous. The purpose of Series F experiments is to test whether the accuracies when tested by the test sentences generated in this research and test sentences taken from Brown corpus will align or not. The Table IV shows the accuracies obtained in Series F experiments on 3541 words when tested by 100 sentences taken from Brown corpus.

TABLE IV. OBSERVED ACCURACIES OF SERIES F EXPERIMENTS

Exp Series	No of Words	No of TS	Number of TS disambiguated correctly						Using PolyWordNet
			Using WordNet						
			Exp 1 Run 1	Exp 1 Run 2	Exp 2 Run 1	Exp 2 Run 2	Exp 3 Run 1	Exp 3 Run 2	Exp 4
Series F	3541	100	36	38	46	38	50	38	62



The results obtained in Series F experiments shows the accuracy of WSD algorithm that uses the PolyWordNet is 62% which is higher than that of the WSD methods that use WordNet. The maximum accuracy obtained by WSD method that uses WordNet is 50%.

TABLE V. ACCURACY (RECALL), PRECISION AND COVERAGE IN SERIES F EXPERIMENTS

Exp Series	WordNet						PolyWordNet
	Exp 1 Run 1	Exp 1 Run 2	Exp 2 Run 1	Exp 2 Run 2	Exp 3 Run 1	Exp 3 Run 2	Exp 4
No of Handled Cases	75	97	97	97	97	98	67
Coverage	75	97	97	97	97	98	67
Precision	48	39.17	47.42	39.17	51.54	38.77	92.53
Accuracy (= Recall)	36	38	46	38	50	38	62

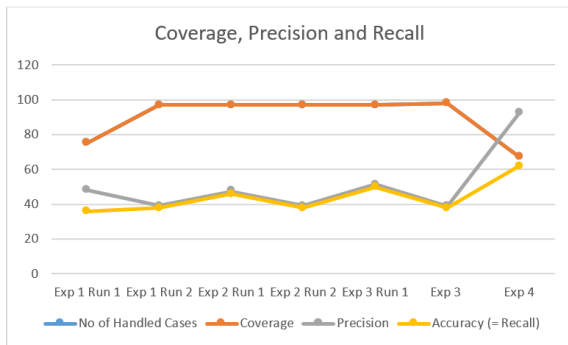


Figure 16. Coverage, precision and accuracy (recall) observed in Series F experiments.

The Table V and Fig. 16 shows the coverage, precision and accuracy obtained in Series F experiments. The WSD algorithm using WordNet has a better coverage 98% than that of the PolyWordNet. The coverage of PolyWordNet is found to be 67% when tested with 100 sentences from Brown corpus. However, the results from experiments shows that WSD algorithm using PolyWordNet has better the precision which is 92.53%. The highest precision of WSD algorithm using WordNet is only 48% with accuracy of 36%.

All these experimental evidence indicates that the word's organization in PolyWordNet is better for word sense disambiguation.

8. VALIDATION OF POLYWORDNET

PolyWordNet is a new lexical database. In contrast with the existing lexical databases, PolyWordNet deals with polysemy words. It organizes the words based on the senses of polysemy words. The PolyWordNet is built from the words taken from WordNet. The words are taken from WordNet and these words are organized based on the principle of PolyWordNet.

WSD method is used as a validation tool for validating word's organization of PolyWordNet as it is done in [23] to validate BalkaNet. The accuracy range of experiments in Series A to E is 35.56% to 60%. In addition, the range of accuracy for Series F experiments is 36% to 50%. This accuracies ranges are aligned with and represent the accuracy ranges of various WSD methods that uses Lesk algorithm and WordNet and tested with standard evaluation exercises like SenseVal. Further, the result from the experiments shows the higher accuracy of WSD method which uses PolyWordNet. These experimental evidences clearly indicate the PolyWordNet is aligned with WordNet and the word's organization in PolyWordNet is acceptable and valid for word sense disambiguation.

Similarly, these accuracies were obtained by testing 2905 test sentences on 3541 data and are found to be aligned with the accuracy ranges of various WSD methods that uses Lesk algorithm and WordNet and tested with standard evaluation exercises like SenseVal. Therefore, these experimental results on other hand clearly indicates that the data used to build PolyWordNet and the test sentences that are used to test the experiments are also valid.

9. POLYWORDNET IS DIFFERENT

PolyWordNet is a new lexical database and deals with the senses of polysemy words and their corresponding related words. It brings the senses of polysemy words and their corresponding related words together forming a cluster of a sense and its related words. The PolyWordNet does not contain any related word common to the senses of the same polysemy word. This resolves the problem of noise information during sense disambiguation.

In a given context, if there is a single related word, it is sufficient to disambiguate the meaning of polysemy word using PolyWordNet. This resolves the problem of insufficient information for sense disambiguation. A brief comparison among the dictionary, WordNet and PolyWordNet is shown in Table IV.



TABLE VI. ACCURACY (RECALL), PRECISION AND COVERAGE IN SERIES F EXPERIMENTS

Comparison Metrics	Dictionary	WordNet	PolyWordNet
Organization of words	Words are organized based on alphabetical order	Words are organized based on synonym set	Words are organized based on the senses of polysemy words
Result of word organization	Words with similar meaning get scattered	Words with similar meaning come together and form a cluster	Senses of polysemy words and their corresponding related words come together and form a cluster
Deals with senses of polysemy words	NO	NO	YES
Deals with related context words	NO	NO	YES
WSD depends on gloss's definition	YES	YES	NO
Produces Noise Information	YES	YES	NO
Provides less(insufficient) distinct inform	YES	YES	NO

PolyWordNet is built based on completely different principle. It deals with and organizes the senses of polysemy words. The Table VI compares the PolyWordNet based on seven metrics which include 1) the way of organizing words, 2) result of the organization of words, 3) whether it deals with polysemy words or not, 4) whether it deals with the related words or not, 5) whether the WSD that uses PolyWordNet depends on gloss's definition or not, 6) whether it produces noise information during disambiguation process or not and 7) whether the information provided for sense disambiguation is sufficient or not.

The PolyWordNet resolves the problems of noise information and insufficient information for sense disambiguation. It organizes the words based on the senses of polysemy words and their corresponding related words. The resulted PolyWordNet is, therefore, especially suitable for word sense disambiguation. There is no any other lexical database that deals with the polysemy words and their related words.

10. IMPLEMENTATION AND SIGNIFICANT

The popular lexical WordNet organizes the words based on synonym sets. That is the words having similar meaning comes together. In addition, this lexical resource relates the words using the different relations such as hypernym, holonym, meronym, hyponym and so on. However, this lexical database does not deal with any relation that connects the senses of polysemy words and related words in a given context. Polysemy words in any natural language are the main cause of sense ambiguity in a context. These polysemy words are creating big problems in any natural language processing tasks whether it is

Machine Translation from one language to another or it is Text summarization.

Fortunately, there exist a relation between the polysemy words and other words in a given context. This means if a given context contain a polysemy word, the context also contains at least one or more related words which can sufficiently disambiguate the sense of the polysemy word in the given context. Thus, there is a strong relationship between a polysemy word and the related words that come together with the context. However, no lexical resources are dealing with this natural relationship of polysemy words and other words in context for word sense disambiguation. If the senses of polysemy words and their corresponding related words are organized by using the relationship that exist naturally in a given context, then such relationships can be used for word sense disambiguation in a given context. This research has utilized this natural relationship among the senses of polysemy words and related words to design a new lexical database called PolyWordNet. PolyWordNet deals with and organizes words based on the relationship between the senses of polysemy words and their corresponding related words. In PolyWordNet, senses of polysemy words are connected with related words.

During the word sense disambiguation of a polysemy word in a given context, the related word(s) from the context are taken and path from those related words are searched. The path of the related words which lead to a sense of the polysemy word is determined. The sense of the polysemy word to which there exists a path from a related word (s) of the context is the correct sense.

The related words are the main key for building PolyWordNet and for word sense disambiguation. In this research, the related words are generated manually which takes lots of time. Therefore, to build a PolyWordNet that contains all possible related words for each senses of polysemy words that exist in a natural language, the related words must be automatically generated. We highly recommend to develop an algorithm for automatic generation of related words in future and self-organizing PolyWordNet utilizing available big corpus. Since PolyWordNet organizes the senses of polysemy words and their corresponding related words, its gives high accuracy for word sense disambiguation. Therefore, it can be used as an intermediate module in every natural language processing task which requires word sense disambiguation. It will be a very useful tool for word sense disambiguation of simple sentences with high accuracy.

The assumptions of this research are: - (1) every context that contains polysemy word also contains related words which can disambiguate the sense of the polysemy word, (2) the PolyWordNet can be used to disambiguate the sense of polysemy words in simple context. The simple context means the context is a simple sentence not a compound sentence and (3) the context should not be ambiguous to human mind. PolyWordNet is a technique to organize the



senses of polysemy words. It is language independent. Therefore, PolyWordNet can be built for any natural language to use for word sense disambiguation in that language.

11. CONCLUSION

A new lexical database- *PolyWordNet* is developed. It organizes the words based on the senses of polysemy words. The PolyWordNet mimics the way how human mind stores and relates the sense of a polysemy word with the related words in a given context. Analogous to human mind, the PolyWordNet organizes the senses of polysemy words with their corresponding related words. Therefore, a sense of a polysemy word and its related words come together and form a cluster. The new WSD method which uses the relations from PolyWordNet, mimics the way how human mind analyzes the sense of a polysemy word with the related words given in the context and finally relates the context with the correct sense of the polysemy word.

The Series A experiments show 96.11% of accuracy of new WSD algorithm which uses the relations from PolyWordNet. The highest accuracy of Simplified Lesk algorithm, which uses the information from WordNet, is found to be 58.33%. In addition, the results of Series F experiments when tested by 100 sentences from Brown corpus shows the accuracy of WSD method using PolyWordNet is 62% with precision 92.53%. This accuracy is higher than the accuracy of the WSD algorithm using WordNet. This proves that the organization of words in PolyWordNet is better than that of the WordNet especially for word sense disambiguation. This is also supported by the results of Series B to E experiments as well.

12. LIMITATIONS AND RECOMMENDATIONS

The main limitation of this research is the number of data that PolyWordNet contains during experiments. The PolyWordNet currently contains altogether 3541 words only. In addition, only 2905 different Test Sentences are used to test experiments. This research highly recommend to further work on automatic generation of related words and thus to develop a self-organizing PolyWordNet.

ACKNOWLEDGMENT

This research is supported by the University Grant Commission (UGC), Nepal. Authors are equally thankful to IDEAS Project (Erasmus Mundus Action 2) for supporting this research.

REFERENCES

- [1] Lesk, M. (1986) Automatic Sense Disambiguation using Machine Readable Dictionaries: How to tell a Pine Cone from an Ice Cream Cone. Proceedings of the 5th Annual International Conference on Systems Documentation. SIGDOC 1986, pages 24-26, New York, NY, USA.
- [2] Fellbaum, Christiane. (2010) WordNet. Springer. Netherlands, Dordrecht.
- [3] Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. (1990) Introduction to WordNet: An on-line lexical database. International Journal of Lexicography, 3(4):235-244.
- [4] Oram, Peter. (2001) WordNet: An electronic lexical database. Christiane Fellbaum (Ed.). Cambridge, MA: MIT Press, 1998. Pp. 423. Applied Psycholinguistics, 22(01):131-134.
- [5] Miller, George A. (1995) WordNet: A lexical database for english. Commun. ACM, 38(11):3941.
- [6] Bhingardive, Sudha and Pushpak Bhattacharyya. (2017) Word Sense Disambiguation Using IndoWordNet. Springer Singapore, Singapore.
- [7] Kolte, Sopan, S.G. Bhirud, and K J Somaiya. (2009) Wordnet: A knowledge source for word sense disambiguation. REVIEW PAPER International Journal of Recent Trends in Engineering: 2.
- [8] Banerjee, Satanjeev and Ted Pedersen. (2002) An adapted Lesk algorithm for word sense disambiguation using WordNet. In Proceedings of the Third Inter-national Conference on Computational Linguistics and Intelligent Text Processing, CICLing 02, pages 136{145, Springer-Verlag, London, UK.
- [9] Fragos, Kostas, Yannis, Maistros, and Christos, Skourlas. (2003) Word sense disambiguation using wordnet relations. In Proc. of the 1st Balkan Conference in Informatics, Thessaloniki.
- [10] Seo, Hee-Cheol, Hoojung, Chung, Hae-Chang, Rim, Sung Hyon, Myaeng, and Soo-Hong Kim. (2004) Unsupervised word sense disambiguation using word-net relatives. Computer Speech & Language, 18(3):253 - 273. Word Sense Disambiguation.
- [11] Montoyo, Andrs, Manuel, Palomar, German, Rigau, and Armando, Surez. (2011) Combining knowledge- and corpus-based word-sense-disambiguation methods. CoRR, abs/1109.2130.
- [12] Roy, Arindam, Sunita, Sarkar, and Bipul Syam, Purkayastha. (2014) Knowledge based approaches to Nepali word sense disambiguation. International Journal on Natural Language Computing (IJNLC), 3(3):51-63.
- [13] Liu, Shuang, Clement, Yu, and Weiyei, Meng. (2005) Word sense disambiguation in queries. In Proceedings of the 14th ACM International Conference on Information and Knowledge Management, CIKM 05, pages 525-532, ACM, New York, NY, USA.
- [14] Sinha, Manish, Reddy, Mahesh K., Pande, Prabhakar, Kashyap, Laxmi and Bhattacharyya, Pushpak. (2004) Hindi word sense disambiguation. In International Symposium on Machine Translation, Natural Language Processing and Translation Support Systems.
- [15] Dhungana, U.R. and Shakya, Subarna. (2014) Word sense disambiguation in Nepali language. In Digital Information and Communication Technology and its Applications (DICTAP), Fourth International Conference, pages 46-50.
- [16] Banerjee, Satanjeev and Pedersen, Ted. (2003) Extended gloss overlaps as a measure of semantic relatedness. In Proceedings of the 18th International Joint Conference on Artificial Intelligence, IJCAI03, pages 805{810, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- [17] Agirre, Eneko and Edmonds, Philip. (2006) Word Sense Disambiguation: Algorithms and Applications (Text, Speech and Language Technology). Springer-Verlag, New York, Inc., Secaucus, NJ, USA.
- [18] Shirai, Kiyooki and Yagi, Tsunekazu. (2004) Learning a robust word sense disambiguation model using hypernyms in definition sentences. In Proceedings of the 20th International Conference on Computational Linguistics, COLING 04, Association for Computational Linguistics, Stroudsburg, PA, USA.
- [19] Dhungana, Udaya Raj and Shakya, Subarna. (2015) Hypernymy in WordNet, its role in WSD, and its limitations. In Computational Intelligence, Communication Systems and Networks (CICSyN), 2015 7th International Conference on, pages 15-19.
- [20] Dhungana, U.R., Shakya, Subarna, Baral, K., and Sharma, B. (2015) Word sense disambiguation using WSD specific WordNet of polysemy words. In Se-mantic Computing (ICSC), IEEE International Conference on, pages 148-152.



- [21] Guthrie, Joe A., Guthrie, Louise, Wilks, Yorick and Aidinejad, Homa. (1991) Subject-dependent Co-occurrence and Word Sense Disambiguation. In pro-ceedings of the 29th Annual Meeting on Association for Computational Linguistics. ACL 91, pp. 146-152, Association for Computational Linguistics, Stroudsburg, PA, USA.
- [22] Kashyap, Laxmi, Salil Rajeev Joshi, and Pushpak Bhattacharyya (2017) In-sights on Hindi WordNet Coming from the IndoWordNet. Springer Singapore, Singapore.
- [23] Tufis, D., Ion, R., & Ide, N. (2004). Word Sense Disambiguation as a Wordnets' Validation Method in Balkanet. LREC.



Udaya Raj Dhungana
Department of Electronics and Computer Engineering, Pulchowk Campus, Tribhuvan University, Lalitpur, Nepal. He received B.E. degree in Computer Engineering from Pokhara University, Pokhara, Nepal in 2005 and M.E. degree in Computer Engineering from Kathmandu University,

Dhulikhel, Nepal in 2011. He is currently pursuing Ph.D. degree in Computer Engineering at Institute of Engineering, Tribhuvan University, Lalitpur, Nepal. He worked as a Ph.D. research exchange student at Department of Computer Science, Darmstadt University of Applied Sciences, Darmstadt, Germany from 2014 to 2015. He has been teaching various computer science courses since 2006 as a teaching assistant and assistant professor in various Universities of Nepal. He is currently working as an assistant professor at Pokhara University since 2013. He is also serving as an ICT director at Pokhara University since 2017. He has been teaching Artificial Intelligence and Prolog as a guest faculty at Darmstadt University of Applied Sciences, Darmstadt, Germany since 2018. He has also served as coordinator of Erasmus Plus scholarship project between Darmstadt University of Applied Sciences, Darmstadt, Germany and Pokhara University, Nepal. His research interest includes the Word Sense Disambiguation, Lexical Database, knowledge representation and expert system, automatic question generation and automatic question answering. He had received Young PhD Fellowship from University Grant Commission (UGC), Nepal and IDEAS Project (Erasmus Mundus Action 2) student exchange scholarship for Ph.D. research. He has also received best paper award at IEEE conference CICSyN 2015, Riga, Latvia.



Prof. Dr. Subarna Shakya
Department of Electronics and Computer Engineering, Pulchowk Campus, Tribhuvan University, Lalitpur, Nepal. Prof. Dr. Subarna Shakya holds Ph.D. in Computer Engineering from Lviv Polytechnic National University, Ukraine. He served as Executive Director at National Information technology Center, Government of Nepal and also

head of Department of Electronics and Computer Engineering, Director of Center for Information Technology and Chairman of Electronics and Computer Engineering Subject Committee, Institute of Engineering, Tribhuvan University. He has also served as coordinator of EURECA (European Research and educational collaboration with Asia) and IDEAS (Innovation and Design for Euro Asian Scholars) project which is financed by the European Commission through the Erasmus Mundus Program project coordinator of LEADER (Links in Europe and Asia for engineering, education, Enterprise and Research exchanges) financed by the European Commission through the Erasmus Mundus. He is visiting research Professor in the University of Rhode Island, USA. He is the advisor member of National Information Technology Committee, Government of Nepal. He is the Life Member of Indian society for mathematical modeling and Computer Simulation, IIT, Kanpur, India, Hon. Fellow Member of Scientific Society of Advanced research and social change, www.ssarsc.org and senior member of science and engineering institute, (SCIEI), www.sciei.org. He is keen interest in research and development of ICT, e-government system, Information security for e-Government system, multimedia system, Computer Systems simulation and modeling, Cloud computing & Security, Natural Language Processing, Energy Efficiency in cloud computing, Information system, computer architecture and software engineering.