A

PROJECT REPORT

ON

WORD SENSE DISAMBIGUATION & WORDNET USING PYTHON

Submitted in partial fulfilment for the requirement of the award of

DEGREE IN

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE & ENGINEERING

Submitted by

Piyush Singh (25500116030)

Moumita Chakraborty(25500116032)

Mouli Paul (25500116033)

Jaysree Gupta (25500116038)

Under the guidance of

Mrs. Rinku Supakar

Head Of The Department Of Computer Science

Dr. Sudhir Chandra Sur Degree Engineering College , Dumdum.

CERTIFICATE

This is to certify that Piyush Singh , Moumita Chakraborty , Mouli Paul and Jaysree Gupta bearing Roll no : 25500116030 , 25500116032 , 25500116033, 25500116038 respectively has carried out their work for the term paper entitled “Word Sense Disambiguation & Wordnet using Python” under my supervision. They have worked sincerely during the duration of the term paper. They have fulfilled all the requirements laid down in the regulations of the B.Tech (4 years) 8th Semester Examination (CS892) of the Department of Computer Science,

Dr. Sudhir Chandra Sur Degree Engineering College , for the session 2019-20.

_____

Signature of the Guide

(PROF. RINKU SUPAKAR)

Head of the Department

Department of Computer Science

Dr. Sudhir Chandra Sur Degree Engineering College , Dumdum

_____

Signature of the HOD

(PROF. RINKU SUPAKAR)

Head of the Department

Department of Computer Science

Dr. Sudhir Chandra Sur Degree Engineering College , Dumdum

_____

Signature Of Examiner

Dr. Sudhir Chandra Sur Degree Engineering College , Dumdum

Acknowledgments

We very gratefully acknowledge our advisor Prof Rinku Supakar for all her guidance at every step . For teaching us how to do research, for imparting so much valuable knowledge and for all her encouragement and words of kindness. We are also very grateful to Prof Rinku Supakar who spent a lot of time to guide us. Thank you for your very valuable advice and for the interest you have shown. We would like to acknowledge the help of the Department of Computer Science of Dr Sudhir Chandra Sur Degree Engineering College.

We wish to express our deep sense of gratitude and indebtedness to Prof. Rinku Supakar, Head of Department of Computer Science of Dr Sudhir Chandra Sur Degree College for providing us the opportunity to utilize the facilities of the Department of Computer Science.

PIYUSH SINGH

(25500116030)

Computer Science & Engineering

DSCSDEC

MOUMITA CHAKRABORTY

(25500116032)

Computer Science & Engineering

DSCSDEC

MOULI PAUL

(25500116033)

Computer Science & Engineering

DSCSDEC

JAYSREE GUPTA

(25500116038)

Computer Science & Engineering

DSCSDEC

Abstract

Natural Language Processing (NLP) is a branch of artificial intelligence that deals with the interaction between computers and humans using the natural language. NLP is used for Machine Translation , Speech Recognition , Sentiment Analysis ,Text Classification. We use NLP for in doing Word Sense Disambiguation.

Words have different meanings based on the context of the word usage in a sentence. Word sense is one of the meanings of a word. Human language is ambiguous,

so that many words can be interpreted in multiple ways depending on the context in which they occur.

Word sense disambiguation (WSD) is the ability to identify the meaning of words in context in a computational manner. WSD is considered an AI complete problem , that is, a task whose solution is at least as hard as the most difficult problems in artificial intelligence.

WSD can be viewed as a classification task: word senses are the classes, and an automatic classification method is used to assign each occurrence of a word to one or more classes based on the evidence from the context and from external knowledge sources. WSD heavily relies on knowledge. Knowledge sources provide data which are essential to associate senses with words. Different approaches for WSD (supervised, unsupervised and Knowledge-based) and evaluation of WSD systems are discussed.

WordNet is a large word database of English Nouns, Adjectives, Adverbs and Verbs. These are grouped into some set of cognitive synonyms, which are called synsets. To use the Wordnet, at first we have to install the NLTK module, then download the WordNet package. WordNet is like a dictionary in that it stores words and meanings. However it differs from traditional ones in many ways. For instance, words in WordNet are arranged semantically instead of alphabetically. Synonymous words are grouped together to form synonym sets, or synsets. Each such synset therefore represents a single distinct sense or concept. Thus, the synset base, alkali represents the sense of any of various water-soluble compounds capable of turning litmus blue and reacting with an acid to form a salt and water.

CONTENT

CHAPTER 1. INTRODUCTION

1.1 Natural Language Processing (NLP) 1

1.2 Applications of NLP 2

1.3 Word Sense Disambiguation 3

1.4 Wordnet 4

1.5 Sense of Wordnet 4

1.6 Parts of speech in Wordnet 4

1.7 Brief History Of Research On Word Sense Disambiguation 5

1.8 Applications of WSD 6

1.9 Objective 6

1.10 Methods of WSD 7

CHAPTER 2. RELATED WORK ON WSD & WORDNET

2.1 Word Sense Disambiguation 8

2.2 BabelNet 9

2.3 Proposed WSD System 10

2.4 Past Work 10

CHAPTER 3. PRESENT WORK

3.1 Lesk Algorithm 11

3.2 Wordnet 13

3.3 Software Required 13

3.4 Hardware Required 13

3.5 Dataset 13

3.6 Input & Output 14

CHAPTER 4. RESULT

4.1 Comparison 18

4.2 Conclusion 18

1. Introduction

1.1 Natural Language Processing (NLP)

Natural language processing (NLP) is a field of artificial intelligence in which computers analyze, understand, and derive meaning from human language in a smart and useful way. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, sentiment analysis, speech recognition, and topic segmentation.

Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken. NLP is a component of artificial intelligence (AI).

The development of NLP applications is challenging because computers traditionally require humans to "speak" to them in a programming language that is precise, unambiguous and highly structured, or through a limited number of clearly enunciated voice commands. Human speech, however, is not always precise -- it is often ambiguous and the linguistic structure can depend on many complex variables, including slang, regional dialects and social context.

There are two main techniques used with natural language processing :

1.**Syntax Analysis**

2. Semantic Analysis.

Syntax Analysis :

Syntax is the arrangement of words in a sentence to make grammatical sense. NLP uses syntax to assess meaning from a language based on grammatical rules. Syntax techniques used include parsing (grammatical analysis for a sentence), word segmentation (which divides a large piece of text to units), sentence breaking (which places sentence boundaries in large texts), morphological segmentation (which divides words into groups) and stemming (which divides words with inflection in them to root forms).

Semantic Analysis :

Semantics involves the use and meaning behind words. NLP applies algorithms to understand the meaning and structure of sentences. Techniques that NLP uses with semantics include word sense disambiguation (which derives meaning of a word based on context), named entity recognition (which determines words that can be categorized into groups), and natural language generation (which will use a database to determine semantics behind words).

Current approaches to NLP are based on deep learning, a type of AI that examines and uses patterns in data to improve a program's understanding. Deep learning models require massive amounts of labeled data to train on and identify relevant correlations, and assembling this kind of big data set is one of the main hurdles to NLP currently.

Earlier approaches to NLP involved a more rules-based approach, where simpler machine learning algorithms were told what words and phrases to look for in text and given specific responses when

those phrases appeared. But deep learning is a more flexible, intuitive approach in which algorithms learn to identify speakers' intent from many examples, almost like how a child would learn human language.

Three tools used commonly for NLP include NLTK, Gensim, and Intel NLP Architect. NTLK, Natural Language Toolkit, is an open source python modules with data sets and tutorials. Gensim is a Python library for topic modeling and document indexing. Intel NLP Architect is also another Python library for deep learning topologies and techniques.

1.2 Applications of NLP

- Machine Translation
- Speech Recognition
- Speech Synthesis
- Information retrieval (IR)
- Information extraction (IE)
- Word sense disambiguation(WSD)
- Parts-of-speech tagging

1.3 Word Sense Disambiguation

One of the first problems that any natural language processing (NLP) system encounters is lexical ambiguity, syntactic or semantic. The resolution of a word's syntactic ambiguity has been solved in language processing by part-of-speech taggers with high levels of accuracy. The problem of resolving semantic ambiguity is generally known as word sense disambiguation (WSD) and has been proved to be more difficult than syntactic disambiguation. Human language is ambiguous, so that many words can be

interpreted in multiple ways depending on the context in which they occur the identification of the specific meaning that a word assumes in context is only apparently simple. Unfortunately, the identification of the specific meaning that a word assumes in context is only apparently simple.

While most of the time humans do not even think about the ambiguities of language, machines need to process unstructured textual information and transform them into data structures which must be analyzed in order to determine the underlying meaning. The computational identification of meaning for words in context is called word sense disambiguation (WSD). Words have multiple meaning based on the context of the word usage in a sentence. Word Sense is one of the meanings of a word .Word Sense Disambiguation (WSD) is the ability to identify the meaning of words in context in a computational manner. WSD is considered as an AI-complete problem, that is, a problem which can be solved only by first resolving all the difficult problems in Artificial Intelligence such as Turing Test.

Example

a)The **bank** will not be accepting cash on Saturdays.

b)The river overflowed the **bank**.

The word **bank** in the first sentence refers to the *commercial (finance) banks*, while in second sentence, it refers to the river **bank** . The ambiguity that arises due to this, is tough for a machine to detect and resolve.

2. WordNet lists five senses for the word pen:

- pen — a writing implement with a point from which ink flows.
- pen — an enclosure for confining livestock.
- playpen, pen — a portable enclosure in which babies may be left to play.
- penitentiary, pen — a correctional institution for those convicted of major crimes.
- pen — female swan.

WSD is one of the central challenges in Natural Language Processing(NLP). Many tasks in NLP require diambiguation. Word Sense Disambiguation is needed in Machine Translation, Information Retrieval , Information Extraction etc. WSD is typically configured as an intermediate task, either as a stand-alone module or properly integrated into an application.

1.4 WordNet

Wordnet is a lexical_database of semantic relations between words . WordNet links words into semantic relations including synonyms, hyponyms, and meronyms. The synonyms are grouped into *synsets* with short definitions and usage examples. WordNet is like a dictionary in that it stores words and meanings. However it differs from traditional ones in many ways. For instance, words in WordNet are arranged semantically instead of alphabetically. Synonymous words are grouped together to form synonym sets, or synsets. Each such synset therefore represents a single distinct sense or concept. Thus, the synset base, alkali represents

the sense of any of various water-soluble compounds capable of turning litmus blue and reacting with an acid to form a salt and water.

1.5 Senses of Wordnet

The senses of Wordnet are classified as follows: -

- Monosemous:- Words with only one sense are said to be Monosemous.
- Polysemous: - The word which has more than one sense is called Polysemous.

1.6 Parts of speech in Wordnet

Wordnet stores information about words that belong to four parts of speech and these are−

(1) Nouns in Wordnet

Noun words have various relations defined in Wordnet for the noun part of speech. These relations are as follows:

- Hypernymy and Hyponymy

These are two most common relations for nouns. These are semantic relationships that connect two synsets if the entity referred to by one is kind of or is a specific example of the entity referred to by other. Specially, if a synset 'A' is kind of a synset, then 'B' is hyponymy of' A' and 'A' is hypernymy of 'B'. The number of hypernym links is equal to the hyponym links for every hypernym links there is a corresponding hyponym links.

(2) Verbs in Wordnet

Verb words have various relations defined in Wordnet for the verb parts of speech. In the senses of a word verb is may be parts of speech of one sense.

(3) Adjectives in Wordnet

The various relations concerning them are−

- Similar- to:

It is defined for adjectives. This is a semantic relationship that links two adjective synsets that are similar in meaning but not close to enough to be put together in the same synset.

- Modifies noun:

This shows the linkage between normal and adjective concepts. It shows certain adjectives can only modify certain nouns. Such adjectives and nouns are linked in Hindi Wordnet by the relation modifies noun.

(4) Adverbs in Wordnet

The various relations concerning them are

- Similar- to
- Modifies noun

These two relations also present on adverbs of the Wordnet.

1.7 Brief History Of Research On Word Sense Disambiguation

WSD is one of the most challenging jobs in the research field of Natural Language Processing. Research work [6] in this domain was started during the late 1940s. In 1949, Zipf proposed his "Law of Meaning" theory. This theory states that there exists a power-law relationship between the more frequent words and the less frequent words. The more frequent words have more senses than the less frequent words. The relationship has been confirmed later for the British National Corpus. In 1950, Kaplan determined that in a particular context two words on either side of an ambiguous word are equivalent to the whole sentence of the context. In 1957, Masterman proposed his theory of finding the actual sense of a word using the headings of the categories present in Roget's International Thesaurus. In 1975 Wilks developed a model on "preference semantics", where the selectional restrictions and a frame-based lexical semantics were used to find the exact sense of an ambiguous word. Rieger and Small in 1979 evolved the idea of individual "word experts". In 1980s there was a remarkable development in the field of WSD research as Large-scale lexical resources and corpora became available during this time. As a result, researchers started using different automatic knowledge extraction procedures (Wilks et al. 1990) parallel with the handcrafting methodologies. In 1986, Lesk proposed his algorithm based on overlaps between the glosses (Dictionary definitions) of the words in a sentence. The maximum number of overlaps represents the desired sense of the ambiguous word. In this approach the Oxford Advanced Learner's Dictionary of Current English (OALD) was used to obtain the dictionary definitions. This approach had shown the way to the other Dictionary-based WSD works. In 1991, Guthrie et al. used the subject codes to disambiguate the exact sense using the Longman Dictionary of Contemporary English (LDOCE). In 1990s, three major developments occurred in the research fields of NLP: online dictionary WordNet [7-13] became available, the statistical methodologies were introduced in this domain, and Senseval began. The invention of WordNet (Miller 1990) brought a revolution in this research field because it was both programmatically accessible and hierarchically organized into word senses called synsets. Today, WordNet is used as an important online sense inventory in WSD research. Statistical and machine learning methods are also successfully used in the sense classification problems. Today, methods that are trained on manually sense-tagged corpora (i.e., supervised learning methods) have become the mainstream approach to WSD. Corpus based Word Sense Disambiguation was first implemented by Brown et al. in 1991. As the data sets, corpuses, online Dictionaries vary language to language all over the world, there was not any bench mark of performance measurement in this domain in the early age. Senseval brought all kind of research works in this domain under a single umbrella. The first Senseval was proposed in 1997 by Resnik andYarowsky. Now, after hosting the three Senseval evaluation exercises, all over the world researchers can share and upgrade their views in this research field.

1.8 Applications of WSD

1.Machine translation

WSD is important for Machine translations. It assists in better understanding of

source language and generation of sentences in target language. It also affects lexical choice depending upon the usage context.
Consider the following two sentences,
I can hear bass sounds.
They like grilled bass.

The occurrences of the word bass in the two sentences clearly denote different meanings: lowfrequency tones and a type of fish, respectively. Here, the process WSD assigns correct meaning to the word bass in the above two sentences as
I can hear bass / low frequency tone sounds.
They like grilled bass / fish.

2.Information extraction and knowledge acquisition

In information extraction and text mining, WSD is required for the accurate analysis of text in many applications. For instance, an intelligence gathering system might need to flag up references to, say, illegal *drugs*, rather than medical *drugs*. More generally, the Semantic Web requires automatic annotation of documents .

3.Text Processing

When words are pronounced in more than one way depending on their meaning is text to speech translation. For example: “lead” can be “in front of” or “type of metal”.

1.9 Objective

☂ The primary objective of the documentation can be summarized as follows:

1. To study WSD.
2. Implementation of WSD using Wordnet.

1.10 Methods Of WSD

1)Knowledge-base Approach:-

It is dictionary base method. Dictionaries have been seen as a likely source of information for use in natural language processing a knowledge base method as a knowledge source has been made like id, word, concept and example. The approach that has been created consists of different fields.
The fields are as follows:

- The first field id of the word. The word id has been taken from the INDOWORDNET.
- The second field is for the word itself. The word has been taken which has more than one meaning.
- The third field is for concept. The concept defines the actual meaning of the word.
- The fourth field is the example. The example is taken with respect to the word which is related to its concept.

2)Supervised Approach:-

Supervised methods are based on the assumption that the context can provide enough evidence on its own to disambiguate words . These supervised methods are subject to a new knowledge acquisition bottleneck since they rely on substantial amounts of manually sense-tagged corpora for training, which are laborious and expensive to create.

3)Semi-supervised Approaches:-

Due to the lack of training corpus, most of the word sense disambiguation algorithms use semi-supervised learning methods. It is because semi-supervised methods use both labelled as well as unlabeled data. These methods require very small amount of annotated text and large amount of plain unannotated text. The technique that is used by semi supervised methods is bootstrapping.

4) Un-supervised Approaches:-

Unsupervised approach unlike supervised approach does not need the hand labeled knowledge of sense information in large scale resources for the disambiguation. It is based on the fact that words having similar senses will have similar surrounding words. Word senses are derived by forming clusters of occurrences of words and the task is to classify the new occurrence to the derived clusters.

2. Related Work On WSD & Wordnet

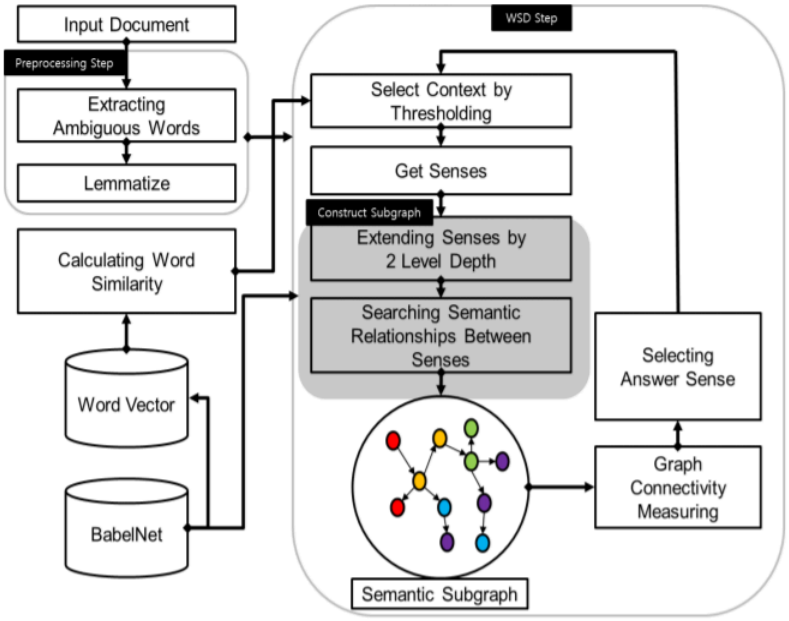
2.1 Word Sense Disambiguation

Recently, the graph-based WSD method become the most popular a method for the knowledge-based WSD (Navigli and Velardi, 2005). The graph-based method selects the answer sense of the ambiguous word based on the semantic structure of LKBs. Generally, the answer sense is chosen from the semantic subgraph that connects the senses of the words in the input document using the semantic relationships defined in LKBs. Navigli and Lapata (2007) built a semantic subgraph of the entire words including senses and then used graph connectivity measures to determine the combination of answer senses. Agirre et al. (2014) suggested a knowledge-based WSD approach used personalized page rank (PPR) over the semantic subgraph. They calculated the relative importance of senses using PPR and the sense with the highest score was chosen as an answer sense. Babelfy (Moro et al., 2014) presented another graph-based approach that jointly selects answer of WSD and entity linking (Xiao et al., 2015). Utilizing the random walk algorithm with a restart, it

extracted a dense subgraph and reweighted the edges of a BabelNet. They iteratively disambiguate words by reconstructing a semantic subgraph at each word. Based on the assumption that word with a minimum sense is an easiest word among the entire ambiguous words, Manion et al. (2014) disambiguated the ambiguous words in order of the number of their senses. Chaplot et al. (2015) maximized the joint probability of whole senses in the context using WordNet and dependency. Tripodi and Pelillo (2017) suggested to apply the idea of the evolutionary game theory to their WSD system. By exploiting the semantic similarity of the words, they formulated WSD as a constraint satisfaction problem and derived it utilizing game theorem tools. In addition to these abovementioned methods, several methods for WSD have been proposed. Chaplot and Salakhutdinov (2018) proposed a topic modeling based WSD approach to ameliorate computational complexity of graph-based WSD. Zhong and Ng (2010) tried to disambiguate words using support vector machine (Suykens and Vandewalle, 1999) with rich linguistic features such as part-of-speech, local collocations and surrounding contextual words. Weissenborn et al. (2015) jointly optimized WSD and entity linking model in an extensible multi-objective optimization. Pasini and Navigli (2017) built a large-scale training corpus for WSD from scratch using Wordnet. Raganato et al. (2017) suggested a supervised WSD approach using bidirectional long short-term memory and attention mechanism. Our WSD method is based on the iterative subgraph reconstruction approach (Manion et al., 2014). However, our WSD approach is crucially different in that it selectively constructs subgraphs by thresholding the contents words of the input document based on the similarity with the ambiguous words.

2.2 BabelNet

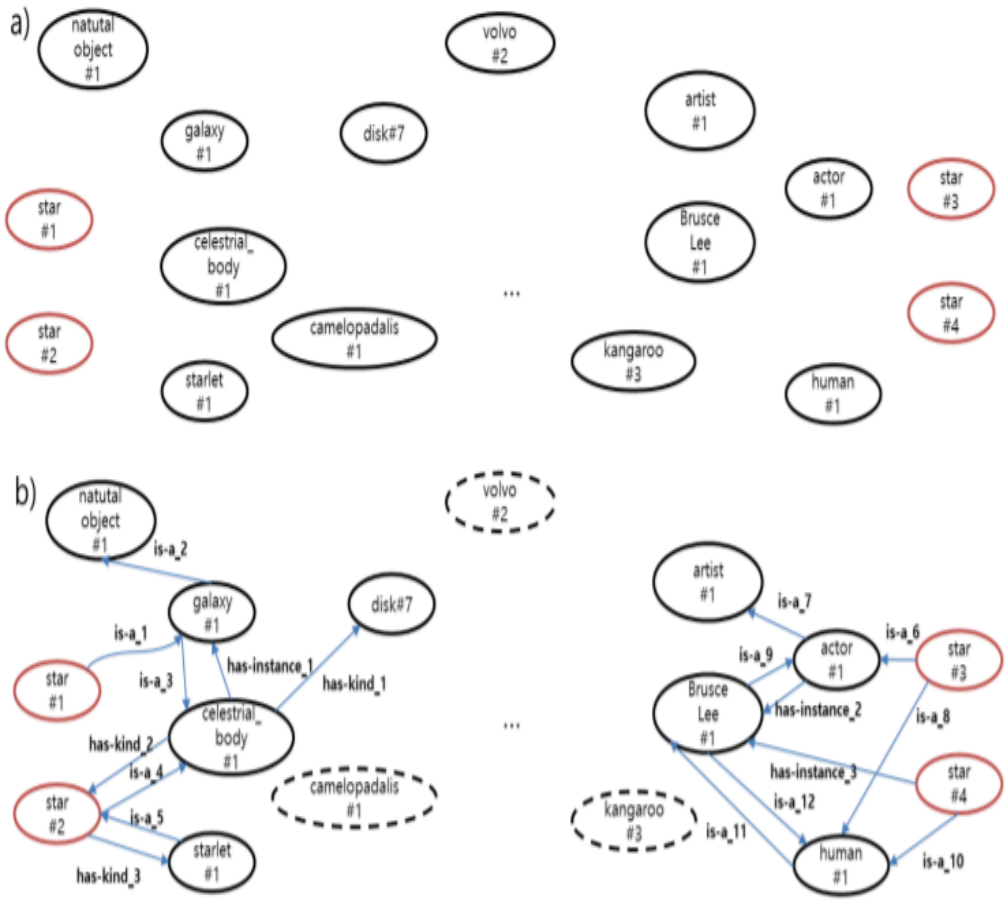
Most unsupervised WSD systems utilize LKBs such as Wordnet1 to obtain a set of possible senses for each ambiguous word. BabelNet2 is a multi-lingual lexicalized semantic network and ontology. It pro-



vides the senses of content words 3 , semantic relationship between the senses and the set of synonyms of the sense. As shown in Figure. 2, BabelNet has a graph structure that consists of nodes and edges. A node indicates the sense of a word and an edge denotes the semantic relationship between the senses. The synonym information for a sense is accessible from Babel synset, which provides multi-lingual synonyms. Semantic relationship between senses contains the relationship defined by both Wikipedia (ACADEMIC_DEGREE, COUNTRY_OF_CITIZENSHIP, DEPICTED_BY, etc) and Wordnet (IS_A, PART_OF, etc). For example, a word ‘Obama’ contains six possible noun senses such as “bn:03330021n: 'Barack Hussein Obama II is an American politician who served as the 44th President of the United States from 2009 to 2017.',” and an adjective sense of “bn:13705874a: 'Of or pertaining to the political figure and 44th president of the United States of America Barack Obama.'.” In addition, ‘Obama’ with a noun sense of ‘bn: 03330021n’ has an 'IS-A' semantic relationship with ‘Human’ with a noun sense of ‘bn: 00044576n’ and has a semantic relationship of 'COUNTRY_OF_CITIZENSHIP' with ‘United States’ with a noun sense of ‘bn: 00003341n.’ As mentioned above, BabelNet provides the senses and their semantic relationships to multi-lingual, it is advantageous to extend a WSD system to other language. This makes many recent studies for WSD systems choosing BableNet as a sense repository (Moro et al., 2014; Manion et al., 2014; Tripodi and Pelillo 2017).

2.3 Proposed WSD System

This section describes our fundamental ideas to improve the performance of WSD and how they are integrated into our WSD system as illustrated in Figure 3. The subsection 3.1 is allocated to introduces a novel word similarity calculation method for the contextual word selection. Next, we explain our new iterative subgraph construction method that combine contextual word selection in the traditional WSD approach.



2.4 Past Work

WordNet is like a dictionary in that it stores words and meanings. However it differs from traditional ones in many ways. For instance, words in WordNet are arranged semantically instead of alphabetically. Synonymous words are grouped together to form synonym sets, or synsets. Each such synset therefore represents a single distinct sense or concept. Thus, the synset base, alkali represents the sense of any of various water-soluble compounds capable of turning litmus blue and reacting with an acid to form a salt and water.

```
from PyDictionary import PyDictionary

dictionary=PyDictionary()

a=dictionary.meaning(word)
```

```
print(a)
```

3. Present Work

3.1 Lesk Algorithm

Lesk algorithm is based on the hypothesis that words used together in text are related to each other and that the relation can be observed in the definitions of the words and their senses. Two (or more) words are disambiguated by finding the pair of dictionary senses with the greatest word overlap in their dictionary definitions. It searches for the shortest path between two words: the second word is iteratively searched among the definitions of every semantic variant of the first word, then among the definitions of every semantic variant of each word in the previous definitions and so on. Finally, the first word is disambiguated by selecting the semantic variant which minimizes the distance from the first to the second word.

In Simplified Lesk algorithm , the correct meaning of each word in a given context is determined individually by locating the sense that overlaps the most between its dictionary definition and the given context. Rather than simultaneously determining the meanings of all words in a given context, this approach tackles each word individually, independent of the meaning of the other words occurring in the same context.

Algorithm

```
function SIMPLIFIED LESK(word,sentence) returns best sense of word
        best-sense <- most frequent sense for word
        max-overlap <- 0
        context <- set of words in sentence

        for each sense in senses of word do
        signature <- set of words in the gloss and examples of sense
        overlap <- COMPUTEOVERLAP (signature,context)

        if overlap > max-overlap then
        max-overlap <- overlap
        best-sense <- sense

        end return (best-sense)
```

Flowchart

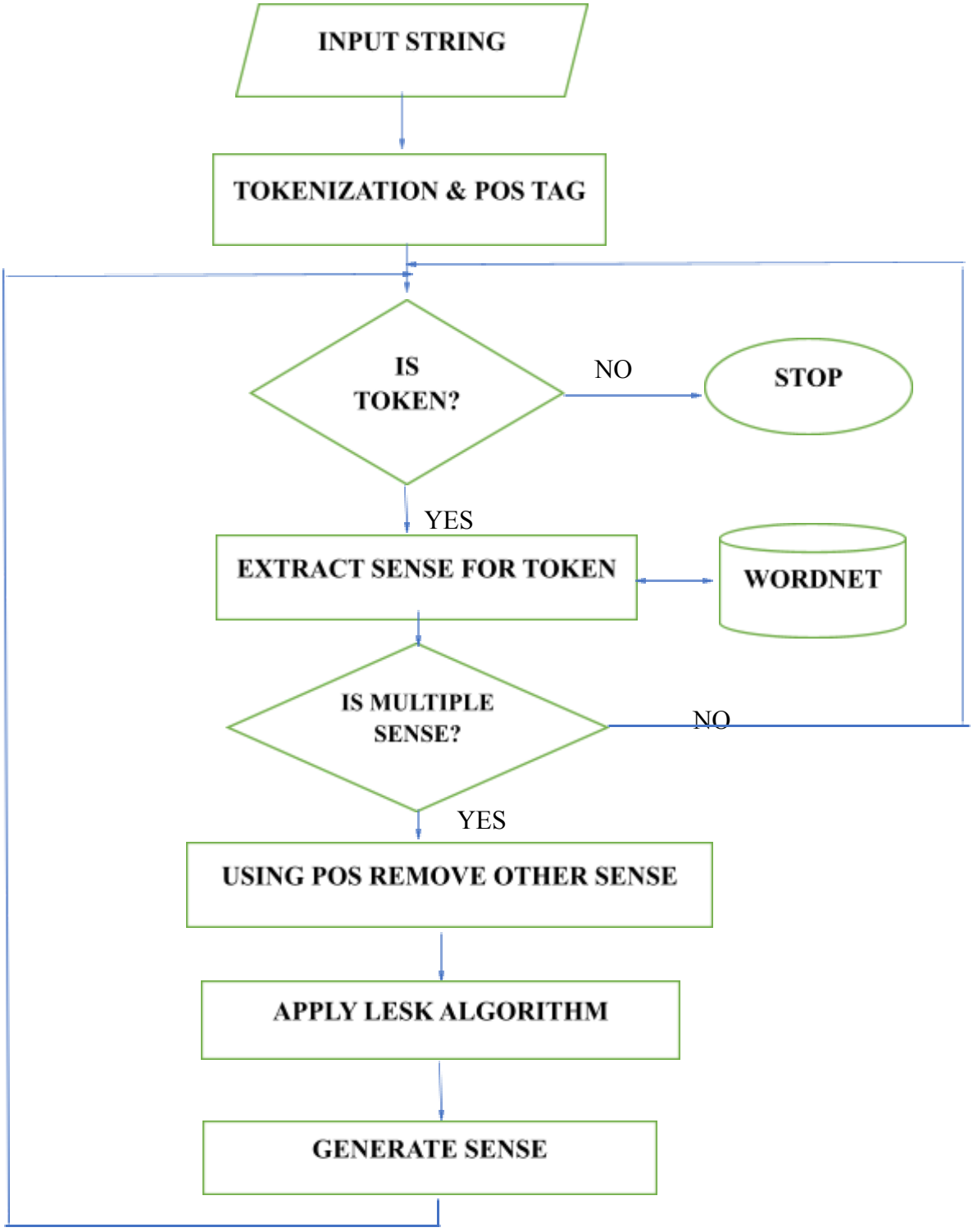


Fig : Flowchart Of WSD utilizing Lesk Algorithm

3.2 Wordnet

WordNet is like a dictionary in that it stores words and meanings. However it differs from traditional ones in many ways. For instance, words in WordNet are arranged semantically instead of alphabetically. Synonymous words are grouped together to form synonym sets, or synsets. Each such synset therefore represents a single distinct sense or concept. Thus, the synset base, alkali represents the sense of any of various water-soluble compounds capable of turning litmus blue and reacting with an acid to form a salt and water.

```
from PyDictionary import PyDictionary

dictionary=PyDictionary()

a=dictionary.meaning(word)

print(a)
```

3.3 Software Required

- PyCharm
- Python GUI Programming(Tkinter)

3.4 Hardware Required

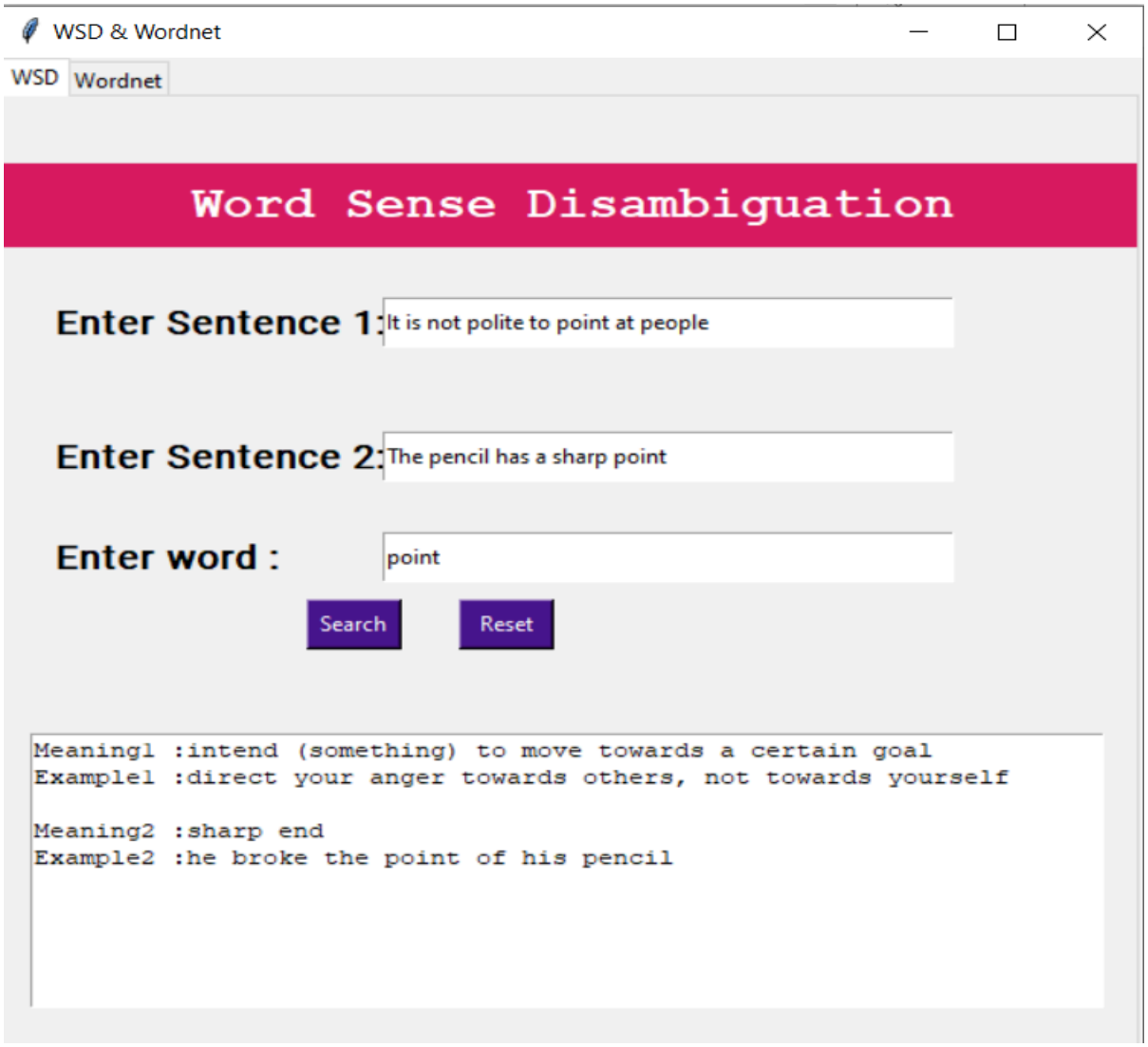
- Processor : Pentium IV or better
- RAM : 128 MB or more
- Hard Disk : 20 GB

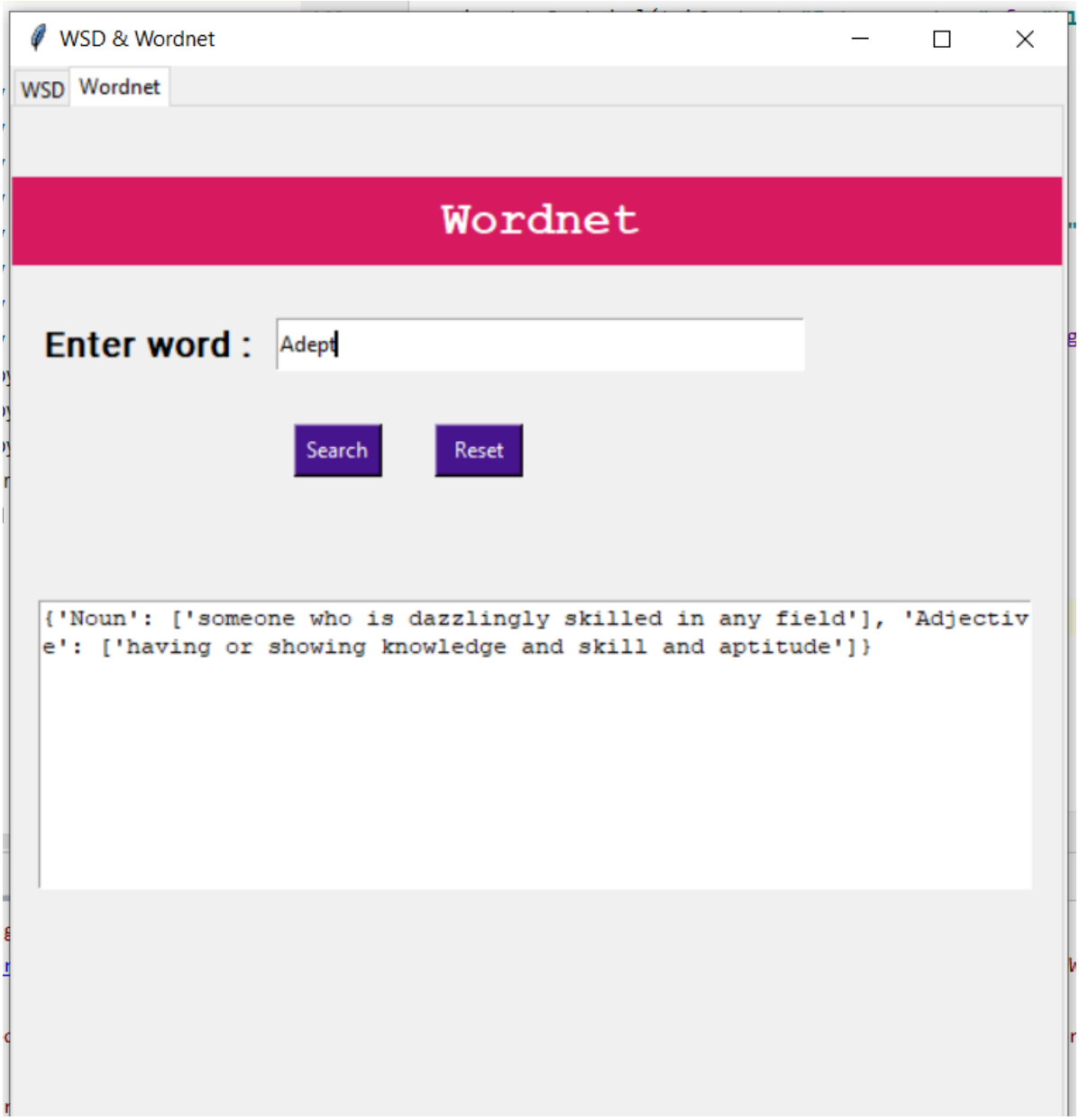
3.5 Dataset

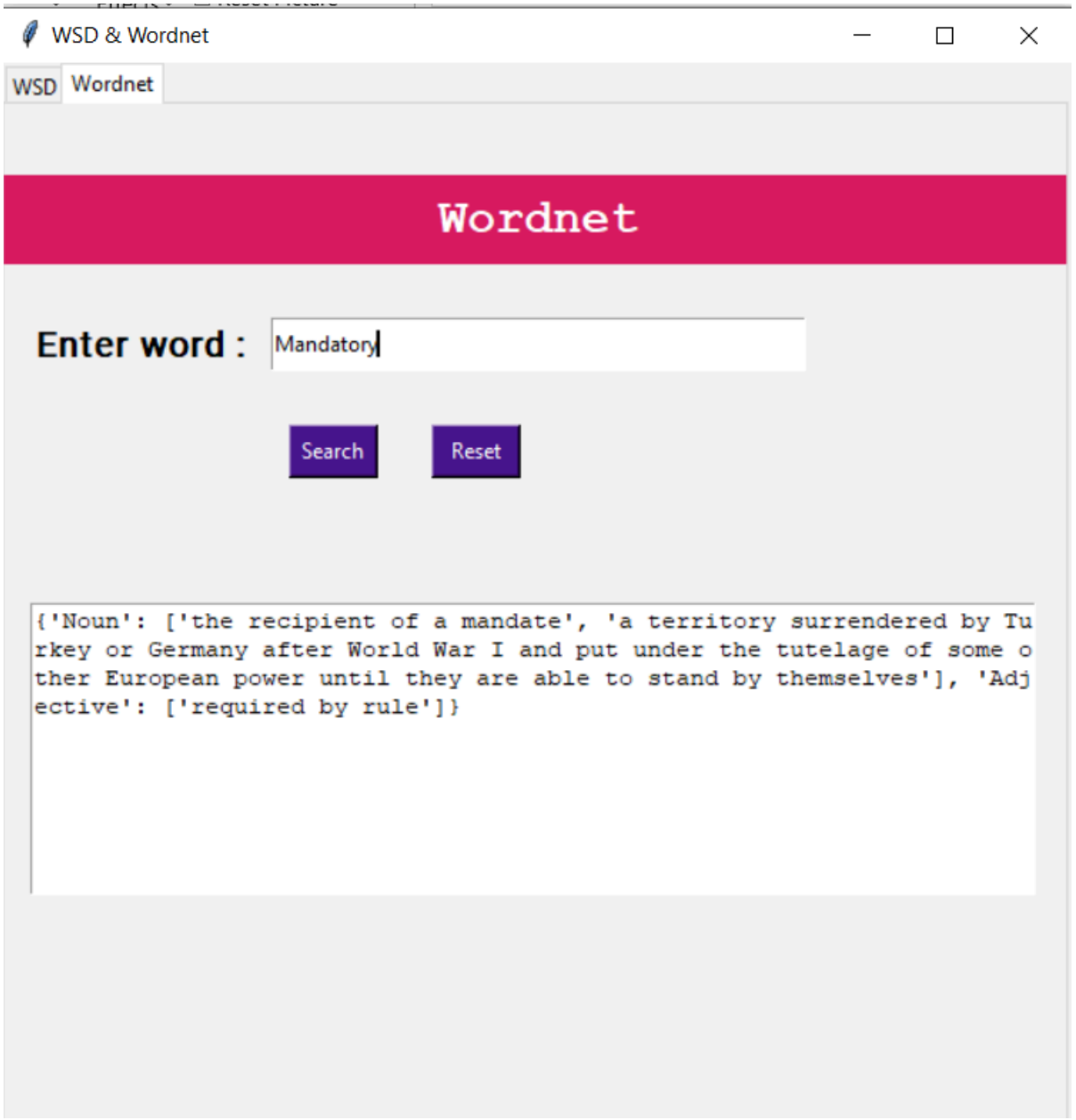
Wordnet is an NLTK corpus reader, a lexical database for English. It can be used to find the meaning of words, synonym or antonym. One can define it as a semantically oriented dictionary of English. It is imported with the following command:

```
from nltk.corpus import wordnet
```

3.6 Input and Output


WSD & Wordnet
WSD
Wordnet
Word Sense Disambiguation
Enter Sentence 1:
It is not polite to point at people
Enter Sentence 2:
The pencil has a sharp point
Enter word :
point
Search
Reset
Meaning1 :intend (something) to move towards a certain goal
Example1 :direct your anger towards others, not towards yourself
Meaning2 :sharp end
Example2 :he broke the point of his pencil


WSD & Wordnet
WSD
Wordnet
Wordnet
Enter word :
Adept
Search
Reset
{'Noun': ['someone who is dazzlingly skilled in any field'], 'Adjectiv
e': ['having or showing knowledge and skill and aptitude']}


WSD & Wordnet
WSD
Wordnet
Wordnet
Enter word :
Mandatory
Search
Reset
{'Noun': ['the recipient of a mandate', 'a territory surrendered by Tu
rkey or Germany after World War I and put under the tutelage of some o
ther European power until they are able to stand by themselves'], 'Adj
ective': ['required by rule']}

4. Result

4.1 Comparison

As the test sets, sense inventories, machine readable dictionaries, knowledge resources which are required for different WSD algorithms are different, each algorithm has some advantage and disadvantage.

Approach	Advantage	Disadvantage
KnowledgeBased	These algorithms give higher Precision.	These algorithms are overlap based, so they suffer from overlap sparsity and performance depends on dictionary definitions.
Supervised	This type of algorithms are better than the two approaches w.r.t. implementation perspective.	These algorithms don't give satisfactory result for resource scarce languages.
Unsupervised	There is no need of any sense inventory and sense annotated corpora in these approaches.	These algorithms are difficult to implement and performance is always inferior to that of other two approaches.

4.2 Conclusion

In this project we try to draw out patterns from the context word and target words in the corpus to train the neural network. The network once trained behaves on a desired output of 57%. The trained data and words acquired from wordnet are too small and so the training must be more vigorous in order to get a better result.

In this section we proposed the semantic and conceptual distance methods to get the correct sense which helped to train the data. The conceptual method draws out a tree structure and finds the shortest path distance from target to context sense. For semantic graph base method we used adjective as modify of the noun and find the shortest distance between the related hypernymy and hyponymy. The data acquired from the result of graph based method was used to train the neural using their occurrences and weight between the context sense and target sense.

References

• Khapra, M.M., Joshi, S. and Bhattacharyya, P. (2011) It Takes Two to Tango: A Bilingual Unsupervised Approach for Estimating Sense Distributions Using Expectation Maximization
https://www.cse.iitb.ac.in/~pb/papers/ijcnlp2011-BilingualEM.pdf

• Bressert, E. (2013) SciPy and NumPy, O'Reilly Media, Inc., USA. Rao, M. S. and Pulari, A.K. (2000) Word Sense Disambiguation Using the Hopfield Model of Neural Networks, Department of Computer and Information Sciences, University of Hyderabad, Hyderabad-5000 046, INDIA

• Khapra, M.M., Shah, S., Kedia, P. and Bhattacharyya Domain-Specific Word Sense Disambiguation combining corpus based and wordnet based parameters, https://www.cse.iitb.ac.in/~pb/papers/gwc2010-english-wsd.pdf

• Khapra, M.M., Joshi, S., Chatterjee, A. and Bhattacharyya, P. (2011) Together We Can: Bilingual Bootstrapping for WSD,
https://www.cse.iitb.ac.in/~pb/papers/acl11-bilingual-bootstrapping.pdf

• Khapra, M.M., Kulkarni, A., Sohoney, S. and Bhattacharyya, P. (2010) All Words Domain Adapted WSD: Finding a Middle Ground between Supervision and Unsupervision,
https://www.cse.iitb.ac.in/~pb/papers/acl2010-da-wsd.pdf Downloaded on

• Khapra, M.M., Joshi, S., Chatterjee, A. and Bhattacharyya, P. (2009) Domain Specific Iterative Word Sense Disambiguation in a Multilingual Setting,
https://www.cse.iitb.ac.in/~pb/papers/icon08-wsd.pdf

• Khapra, M.M., Joshi, S., Chatterjee, A. and Bhattacharyya, P. (2009) Projecting Parameters for Multilingual Word Sense Disambiguation,
https://www.cse.iitb.ac.in/~pb/papers/emnlp09-wsd.pdf