ISSN No. 0976-5697

Volume 5, No. 6, July-August 2014

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Word Sense Disambiguation using Phases of Compiler

Romika Yadav[#1], Garima[#2]
[#]Indira Gandhi University Rewari, India

Amit Seharawat[*3]
[*]GITM, Bilaspur Gurgaon, India

Abstract **-** Word Sense Disambiguation is a known problem which can be accurately stated and which is assumed to have an objective and provablesolution. Many solutions are exist to enhance words sense disambiguation such asMINION, Association Rule, Syntactic Analysis etc. In this paper we are proposed a model to enhance word sense disambiguation by using phases of compiler. The meaning of a word (ambiguous) is provided in a given perspectiveof a sentence. First we make a sentence using that word which causes an ambiguity. Then we construct a syntax tree for the sentence and chunking is performed. It is proved to resolve the ambiguity.

Key Words: Word Sense Disambiguation, Natural Language Processing, Syntax Tree, Ambiguous, Compiler.

I. INTRODUCTION

Word Sense Disambiguation is the problem of assigning a sense to ambiguous word, using its perspective. Word Sense Disambiguation is the problem of determining which "sense" (meaning) of a word is activated by the use of the word in a particular context, a process which appear to be largely unaware in people. WSD is a natural classification problem. Given a word and its possible sense, as defined by a dictionary, classify an occurrence of the word in context into one or more of its sense classes. The feature of the context (such as neighboring of the word) provides the evidence for classification.A famous example is to determine the sense of **Can** in the following passage. Two senses for the word **Can**:

Can: I **can** do it.

Can: A **can** of soda has 100 calories.

Machine translation is the original and most oblivious application for WSD but it has actually been considered in almost every application of language technology, including information retrieval, lexicography, knowledge mining/acquit ion and semantic interpretation and in becoming increasingly important in new research areas such as bioinformatics and semantic web.

There are 4 conventional approaches to Enhance Word Sense Disambiguation

a) ***Dictionary and Knowledge base Methods:*** These rely primarily on dictionaries, thesauri and lexical knowledge bases without using any corpus evidence.

b) ***Supervised Methods:*** These make use of sense annotated corpora to train from.

c) ***Semi Supervised Methods:*** These make use of a secondary source of knowledge such as a small annotated corpus as seed data in a bootstrapping process, or a word aligned bilingual corpus.

d) ***Unsupervised Methods:*** These almost completely external information and work directly from raw un annotated corpora. These methods are known under the name of WSD.

A. Phases of compiler:

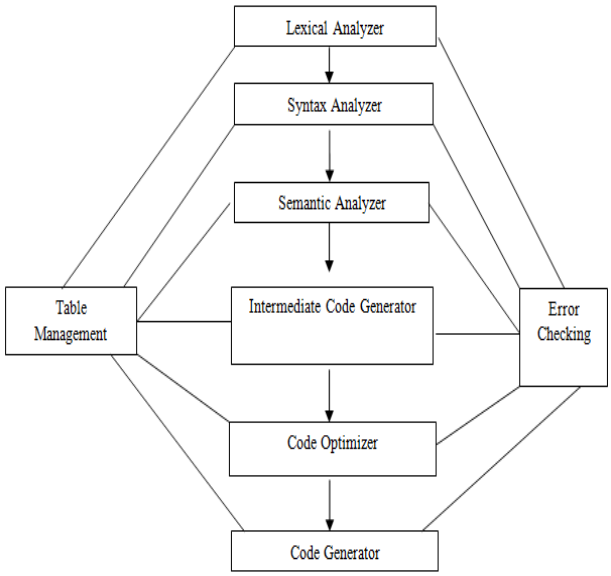

Figure 1. Phases of Compiler

Compiler operates in phases Fig. 1each phase transforms the sentence from one representation to another phases are:

a. ***Symbol table management:*** It is a data structure containing a record for each symbols(word) , with fields for the attributes of the identifier.Records the words used in the sentence and collect information about the word such as its categories, its scope etc.

b. ***Error detecting:*** Each phase encounters errors.

c. ***Analysis phase:*** It involves: lexical, syntax and semantic phase.

d. ***Lexical analysis:*** This phase records the character in the sentence and group them into a stream of tokens in which each token represent a logically cohesive sequence of characters such as categories, keywords, punctuation character etc.

e. ***Syntax analysis:*** Syntax analysis imposes a hierarchical structure on the token stream. This structure is called syntax tree.

f. ***Semantic analysis****:* This phase checks the sentence for semantic error and gather type information for the subsequent code generation phase.The hierarchical structure determined by the syntax analysis phase to identify the symbols.

g. ***Intermediate code generation:*** Syntax and semantic analysis generate a explicit intermediate representation of the sentence. It should have two properties: It should easy to produce, easy to translate into target program. In this, chunking of the words is performed.

h. ***Code optimization:*** These attempts to improve the intermediate code, so that faster running machine code will result.

i. ***Code generation:*** Final phase of the compiler is the generation of the target sentence(meaning).

II. LITERATURE SURVEY

Provide an approach to improve a word sense disambiguation using MINION that is constraint solver. The correctly aligned words are collect from the MINION and then rules are formed using the CLIPS language, finally get the correct sense of a word. Giving the correct and simple sense of a word in a given context[1].Provide an approach to intensification WSD using topic features that comprise of latent dirichlet allocation (LDA) algorithm on the data. This paper integrated the features in the modification in Naive Bayes network like syntactic patterns, part of speech words, and single word context. This improved method achieved development and also more exactness over the simple Naive Bayes network [2]. Provide an approach to select an feature word by two methods that is dependency based and window based methods.

The proposed method in this paper selecting the features accurately with the help of syntactic parsing.Dependency parsing and phrase structure parsing are the kind of syntactic parsing that is used in this paper [3].Provide an approach to get the correct sense of word in a given context by using the association rule, corpus-based methods, and knowledge based methods. This paper giving an approach to construct a database for word sense disambiguation by using the association rule, which can be used to mine the correct sense of a ambiguous word [4].Provide an approach to find then correct sense of a word in a given perspective in which the word occurs. The repository sense can derive from WordNet (Computation lexicon), Dictionary (Machine readable) and a thesaurus [5]. Provide an approach improving the Influence of Subjectivity on Word Sense Disambiguation onRelated Opinion Analysis. It provides a combination of Subjectivity Word Sense Disambiguation (SWSD) into the related opinion analysis for increase the performance [6].

III. PROBLEM DEFINITION

This paper word sense disambiguation using phases of compiler gives another solution to disambiguate a word. It is the aim that the many approaches are also exist to get the correct sense of a ambiguous word, we provide an accurate and simple approach to get the meaning of a word using phases of compiler.

IV. PROPOSED WORK

In this paper, we are going to present a sentence which has ambiguous meaning. In an example given below, we are taking a word **Matter** that has two meanings द्रव्य. and विषय. , according to the sentence to get the correct meaning of a word, first we have to parse the sentence according to the rules of syntax tree. The syntax tree is to be presented in fig. 2.

Sentence = Matter is a substance that has rest mass.

Sense1: Matter refers to द्रव्य.

Sense2: Sense: Matter refers to विषय.

Parse Tree

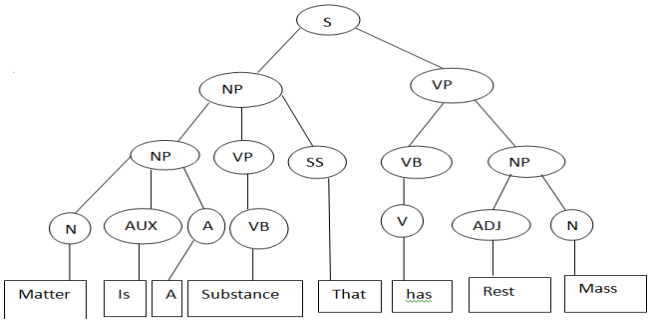

Figure. 2

The figure 2 explains the categories of a sentence according to the Hindi meaning of a word. Here ‘Matter’ is noun, ‘Is’ an auxiliary,’A’ is an article, ‘Substance’ is a verb,’That’ is itself a sentence so it is denoted by ‘SS’, ‘Has’ is a verb,’Rest’ is an adjective,’Mass’ is a noun. The Hindi meaning of these words is taken from the conventional dictionaries.

Table 1

PHASES	OUTPUT	EXAMPLE
Sentence	Source program	Matter is a substance that has rest mass.
Lexical Analysis	Tokens	‘Matter’, ’Is’, ’A’, ’Substance’, ’That’, ’Has’, ’Rest’, ’Mass’.
Syntax Analysis	Parse Tree	Refer Tree (Fig. 2)
Semantic Analysis	Parse Tree	Parsing Category of a sentence (Refer Fig. 2)
Intermediate code generation	Chunking	SS=that S1=adj+N S2=V+S1 S3=N+Aux+A S4=VB+S3 S5=S4+SS S6=S5+S2
Code optimization	Intermediate code	S1=N+Aux+A+VB S2=V+Adj+N S3=S1+SS+S2
Code generation	Generate the output	Meaning is generated

Above table 1 calculating the meaning of a word in a given context. After generation of syntax tree, chunking is to be performed.The chunking is performed in two ways, intermediate code generation and code optimization. In intermediate code generation the values are to be stored in another variable (S1,S2,S3,S4,S5,S6). In code optimization, the output of intermediate code generation is optimized to three variables (S1,S2,S3). Finally from the chunking the correct meaning of a word according to given context is obtained.

V. CONCLUSION AND FUTURE WORK

This paper improve the problem of Word Sense Disambiguation. This problem is very easilysolved by the phases of compiler. This model remove the am understand the proper meaning of each word that causes ambiguity. Every phase provide the efficient and proper meaning of the sentence. The syntax tree of the sentence was exploited extensively.Future work of this paper is that how to improve the efficiency and accuracy of an ambiguous word.

VI. REFERENCES

[1]. Romika Yadav and Rashmi Manker,"Improvement of Word Sense Disambiguation Using MINION," International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 2,Issue 3, March -2013.

[2]. Jun Fu Cai, Wee Sun Lee, Yee Whye The, "Improving Word Sense Disambiguation Using Topic Features,"Proceedings in Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1015–1023, Prague, June 2007.

[3]. Wenpeng LU, Heyan HUANG, and Chaoyong ZHU, "Feature Words Selection for Knowledge-based Word Sense Disambiguation with Syntactic Parsing," PRZEGLĄD ELEKTROTECHNICZNY (Electrical Review), ISSN 0033-2097, R. 88 NR 1b/2012.

[4]. Ajaykumar,Samitkumar, " Word sense disambiguation using Association Rules: A review," IJIRS, ISSN 2319-9725,2013.

[5]. Deepesh Kumar Kimtani, JyotirmayeeChoudhury, AlokChakrabarty, "Improvement in Word Sense Disambiguation by introducing enhancements in English WordNet Structure," International Journal on Computer Science and Engineering (IJCSE), ISSN: 0975-3397, Vol. 4, 07 July 2012.

[6]. CemAkkaya, JanyceWiebe, Alexander , " Improving the Impact of Subjectivity Word Sense Disambiguation on Contextual Opinion Analysis," Proceedings of the Fifteenth Conference on Computational Natural Language Learning,pages 87–96, Portland, Oregon, USA, 23–24 June 2011.