Knowledge Based Approaches to Nepali Word Sense Disambiguation

Arindam Roy[1],Sunita Sarkar[2]and Bipul Syam Purkayastha[3]

[1]Department of Computer Science, Assam University,Silchar
[2]Department of Computer Science,Assam University,Silchar
[3]Department of Computer Science, Assam University,Silchar

ABSTRACT

A word may have multiple senses and the challenge is to find out which particular sense is appropriate in a given context. Word sense disambiguation(WSD) resolves this ambiguity by finding out which particular sense of a word is appropriate in a given context. WSD is of critical importance in the areas of machine translation, information retrieval, speech processing etc. In this paper we present some approaches to Word sense disambiguation in Nepali using Nepali WordNet. These approaches are overlap based approach and conceptual distance and semantic graph based approach which falls under Knowledge based approach. Conceptual distance and semantic graph distance are used as a measures to score our WSD algorithm.

KEYWORDS

WordNet, WSD, conceptual distance, knowledge based approach, semantic graph,overlap based approach.

1. INTRODUCTION

All natural languages have ambiguous words which need to be disambiguated so that the appropriate sense of a word in a given context can be identified.WSD is used to identify the appropriate sense of the target word in a given context. The surrounding words of the target word in a sentence provide the context for the target word and this context provides consistent clue as regards the appropriate sense of the target word. In knowledge based approach to WSD, a machine readable lexical database in the form of Nepali Wordnet [23] has been used. The Nepali WordNet has been developed at Assam University , Silchar as part of a Consortium Project headed by IIT, Bombay with a generous grant from Department Of Information Technology, Ministry of Communications and Information Technology, India. It is a machine readable lexical database for the Nepali language along the lines of the famous English Wordnet[3] and the Hindi Wordnet[4].

2. ROAD MAP

The roadmap of the paper is as follows: Section 3 is on literature survey.Section 4 provides a description of Nepali WordNet. Section 5 and its subsections presents the framework and methodology for knowledge based approach to Nepali Word sense disambiguation. Section 6 describes experimental results and Section 7 winds up the discussion by presenting the conclusions .

10.5121/ijnlc.2014.3305

3. LITERATURE REVIEW

Major WSD approaches proposed till date can be broadly classified as *Knowledge Based Approaches and Machine Learning Based Approaches*.

Knowledge-based WSD is based on lexical resources like dictionaries, thesauri, and corpora where Machine-readable dictionaries (MRDs) are the primary source of acquisition of data. There are various Knowledge based approaches such as WSD using Selectional Preferences [11],Lesk's algorithm [7],Walker's algorithm[10], WSD using conceptual density[1] and WSD using Random Walk Algorithm[12].

Lesk was the first to use dictionary definitions to disambiguate words. To automatically decide which sense of a word is intended, the Lesk algorithm counts overlapping content words in the sense definitions of the target word and in the definitions of context words occurring nearby. Overlap based algorithms typically suffer from sparse overlap, as dictionary definitions are generally small in length. Pedersen T. and Banerjee S.[19] provide an extension to Lesk algorithm

Another knowledge based approach proposed by Agirre Eneko & German Rigau[1] uses the conceptual distance between the senses of the context words and the sense of the target word as a measure for disambiguation. They proposed a formula for conceptual distance which is inversely proportional to the length of the path between two synsets in the wordnet graph and directly proportional to the depth of the two synsets in the WordNet hierarchy.

Corpus based approaches use sense-tagged corpus as the basis for performing WSD. Weiss [20] demonstrated that disambiguation rules can be learnt from a manually sense-tagged corpus using a small study of five words, a training set of 20 sentences for each word, and 30 test sentences for each word. Edward et.al.[21] worked on 1800 ambiguous words from a corpus of a half-million words, with concordance as the basis for manual creation of disambiguation rules ("word tests") for each sense of the 1800 words. Black [22] developed a model based on decision trees using a corpus of 22 million tokens, after manually sense-tagging approximately 2000 concordant lines for five test words.

The study of machine learning based algorithms (supervised as well as unsupervised) suggested that extracting "sense definitions" or "usage patterns" from corpora helps in improving the accuracy of WSD. However, most supervised algorithms which perform very well are not general purpose WSD systems, but word specific classifiers (for example, WSD using SVM [5], Exemplar based WSD [13]) and Yarowsky's decision list algorithm[15].Some of the finer senses of a word, at times, cannot be distinguished by these algorithms. Finally, the requirement of a large training corpus renders these algorithms unsuitable for resource poor languages of which Indian languages are examples.

Hybrid approaches like WSD using Structural Semantic Interconnections [8] use combinations of more than one knowledge sources (wordnet as well as a small amount of tagged corpora). This allows them to access important information encoded in wordnet as well as draw syntactic generalizations from minimally tagged corpora.

4.NEPALI WORDNET

Nepali WordNet is a system for bringing together different lexical and semantic relations between the Nepali words. It organizes the lexical information in terms of word meanings and can be termed as a lexicon based on psycholinguistic principles. The design of the Nepali WordNet is based on the principle of "expansion" from the Hindi WordNet and English WordNet. This principle was first proposed within the Euro WordNet project[22]. Thereafter it has been used by a number of WordNet development teams for the creation of new Wordnets. Examples include the WordNets for Spanish, French, Hungarian language etc. In the Expansion Approach, synsets of a preexisting WordNet are understood by the lexicographer and the corresponding target language synsets expressing the same sense are created.

4.1 Features Of Nepali WordNet

In Nepali WordNet, the words are grouped together according to their similarity of meanings. Two words that can be interchanged in a context are synonymous in that context. For each word there is a synonym set, or synset, in the Nepali WordNet, representing one lexical concept. This is done to remove ambiguity in cases where a single word has multiple meanings. Synsets are the basic building blocks of WordNet. The Nepali WordNet deals with the content words, or open class category of words. Thus, the Nepali WordNet contains the following category of words- Noun, Verb, Adjective and Adverb. Each entry in the Nepali Synset consist of the following elements:-

ID: The synset identifier.
POS: The part of speech of the word.
CONCEPT: It explains the concept represented by the synset.

For example, "यस्तो कुरा वा काम जसले कसैको मान वा प्रतिष्ठा कम गराउँछ" (*yastokuraa waa kaam jasle kasaiko maan waa pratishTha kam garaaũcha*) explains the concept of insult as some saying or deed which diminishes somebody's reputation.

EXAMPLE: It gives the usage of the words of the synsets in the sentence. In general, the words in a synset are replaceable in the sentence. For example: "हामीले कसैलाई पिन अपमान गर्नुहुँदैन (*haameele kasailaaee pani apmaan garnuhũdain*) gives the usage for the words in the synset of 'अपमान', '*apmaan*' representing insult as something that should not be done to anybody.

5. APPROACHES TO NEPALI WORD SENSE DISAMBIGUATION

Broadly two approaches to Nepali WSD are discussed in this paper. These are the overlap based approach and conceptual distance and semantic graph based approach which together falls under knowledge based approach.

5.1 The Overlap Based Approach

The overlap based approach consists of the following :-

i) Preprocessing phase: The preprocessing phase consists of the following steps:

a) Tokenizing: Tokenizer parses the Nepali sentence into words based on the space between words.

b) Context Selection: This module uses the words of the sentence itself as context, including target words, but stop-words like conjunctions, articles, pronouns etc are discarded. Let this collection be w.

c) Finding senses of the target word: This module finds all the possible senses of target word with the help of the Nepali WordNet and forms a collection of words from:

- Synonyms in the synsets
- Glosses of the synsets
- Example sentences of the synsets
- Hypernyms
- Glosses of Hypernyms
- Example Sentences of Hypernyms (upto 2 levels)

Let this collection be called as c_i where i=1,2,…,n.

ii)Determining the Winnner Sense:- The maximum number of overlapping words i.e.words in the context of the target word common to c_i(w in c_i) is determined by the module. The collection c_i which has the maximum number of overlapping words is the winner sense.

IIustrative Example: The Nepali WSD module takes a sentence as input . The target word, context words and the hypernyms of target words and context words are stored as xml files. The target word is enclosed within<>.

For example let us consider this sentence:-

संत कबीर ज्ञान के सागर □□□ <सागर>

The system removes the pronouns and short words(2 or less) from the sentence. Then it finds the senses of the target word <सागर>

Sense 1:- सागर, समुद्र, सिंधु, अंबुधि

Gloss:- खाने पानीको त्यो विशाल राशि जुनचाहिँ चारैतिरबाट पृथ्वीको स्थल भागले घेरिएको छ(*a division of an ocean or a large body of salt water partially enclosed by land*)

Example:- "समुद्र रत्नहरूको खानी हो / रामले वानर सेनाका सहायताले समुद्रमा सेतुको निर्माण गरेका थिए"

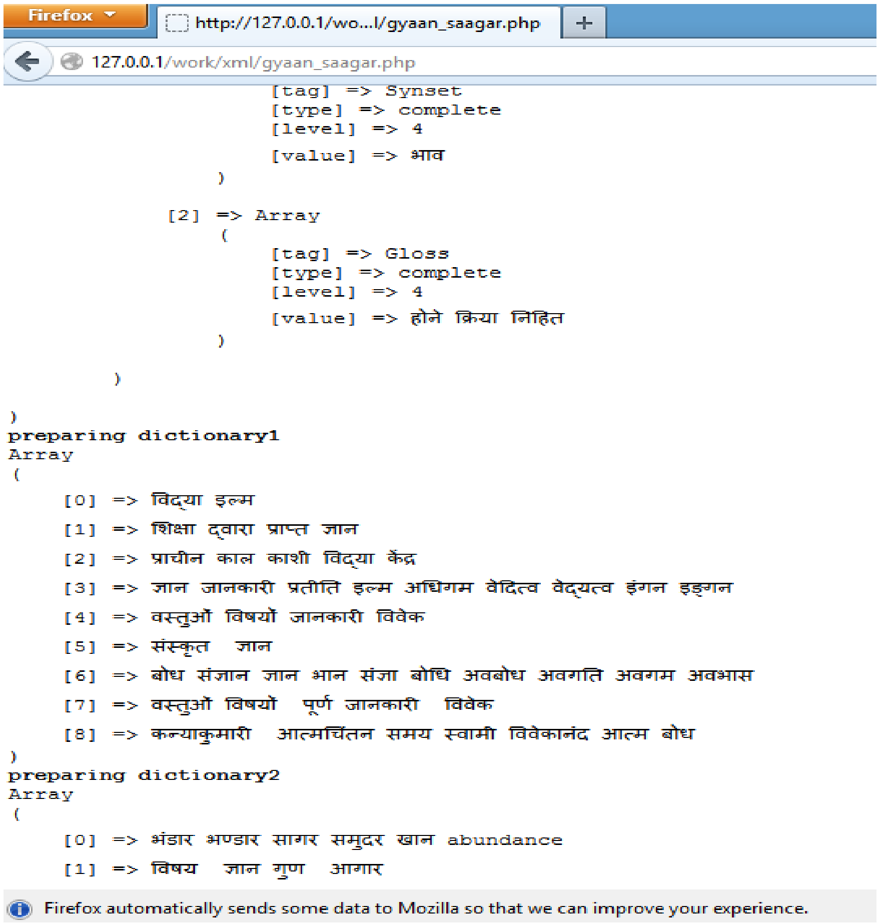
Sense 2:- भंडार भण्डार सागर समुदर खान

Gloss2:- □□□□ □□□□□□□ □□□□□ □□ □□□ □□□□□ □□□□ □□□□ □□□□□□□(*an abundant source of information)*

Example2:- संत कबीर ज्ञान के सागर □□□

The overlap or the word common between the senses of the target word and context word is ज्ञान.Therefore the winner sense is the sense 2 of सागर which means abundance in English

Some screen shots of the execution of the above example are shown :-

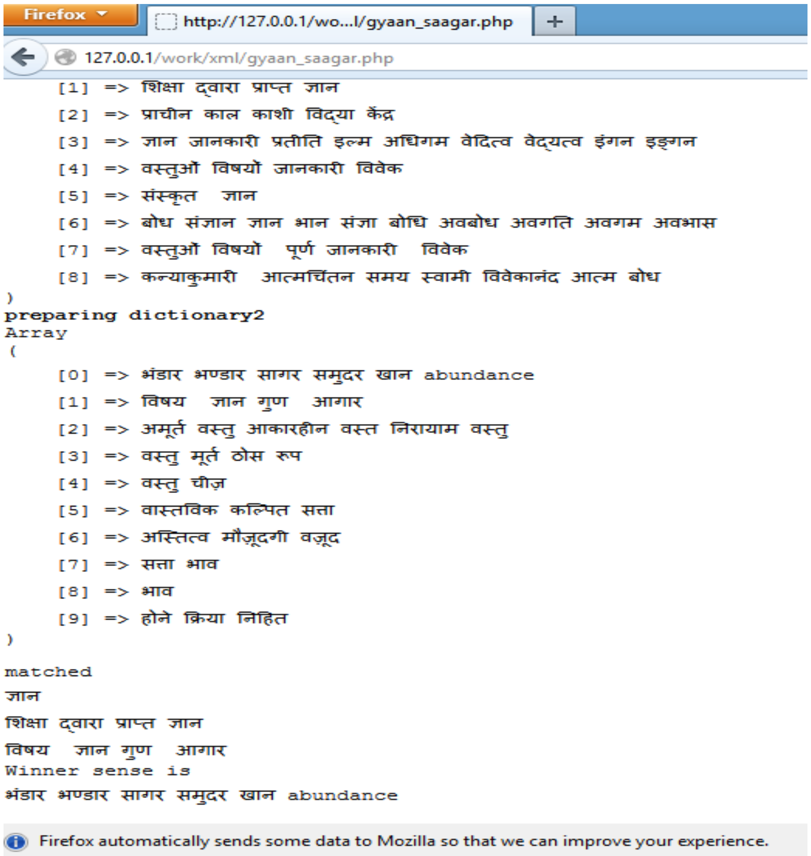


Firefox
http://127.0.0.1/wo...l/gyaan_saagar.php
127.0.0.1/work/xml/gyaan_saagar.php

```
                    [tag] => Synset
                    [type] => complete
                    [level] => 4
                    [value] => भाव
                )

            [2] => Array
                (
                    [tag] => Gloss
                    [type] => complete
                    [level] => 4
                    [value] => होने क्रिया निहित
                )

        )

)
preparing dictionary1
Array
(
    [0] => विद्या इल्म
    [1] => शिक्षा द्वारा प्राप्त ज्ञान
    [2] => प्राचीन काल काशी विद्या केंद्र
    [3] => ज्ञान जानकारी प्रतीति इल्म अधिगम वेदित्व वेद्यत्व इंगन इङ्गन
    [4] => वस्तुओं विषयों जानकारी विवेक
    [5] => संस्कृत  ज्ञान
    [6] => बोध संज्ञान ज्ञान भान संज्ञा बोधि अवबोध अवगति अवगम अवभास
    [7] => वस्तुओं विषयों  पूर्ण जानकारी  विवेक
    [8] => कन्याकुमारी  आत्मचिंतन समय स्वामी विवेकानंद आत्म बोध
)
preparing dictionary2
Array
(
    [0] => भंडार भण्डार सागर समुदर खान abundance
    [1] => विषय  ज्ञान गुण  आगार
```

Firefox automatically sends some data to Mozilla so that we can improve your experience.

Fig1:A screenshot showing example of the overlap based approach


Firefox
http://127.0.0.1/wo...l/gyaan_saagar.php
+
127.0.0.1/work/xml/gyaan_saagar.php

```
    [1] => शिक्षा द्वारा प्राप्त ज्ञान
    [2] => प्राचीन काल काशी विद्या केंद्र
    [3] => ज्ञान जानकारी प्रतीति इल्म अधिगम वेदित्व वेद्यत्व इंगन इङ्गन
    [4] => वस्तुओं विषयों जानकारी विवेक
    [5] => संस्कृत  ज्ञान
    [6] => बोध संज्ञान ज्ञान भान संज्ञा बोधि अवबोध अवगति अवगम अवभास
    [7] => वस्तुओं विषयों  पूर्ण जानकारी  विवेक
    [8] => कन्याकुमारी  आत्मचिंतन समय स्वामी विवेकानंद आत्म बोध
)
preparing dictionary2
Array
(
    [0] => भंडार भण्डार सागर समुदर खान abundance
    [1] => विषय  ज्ञान गुण  आगार
    [2] => अमूर्त वस्तु आकारहीन वस्त निरायाम वस्तु
    [3] => वस्तु मूर्त ठोस रूप
    [4] => वस्तु चीज़
    [5] => वास्तविक कल्पित सत्ता
    [6] => अस्तित्व मौज़ूदगी वज़ूद
    [7] => सत्ता भाव
    [8] => भाव
    [9] => होने क्रिया निहित
)

matched
ज्ञान
शिक्षा द्वारा प्राप्त ज्ञान
विषय  ज्ञान गुण  आगार
Winner sense is
भंडार भण्डार सागर समुदर खान abundance
```

Firefox automatically sends some data to Mozilla so that we can improve your experience.

Fig2:Another screenshot showing example of the overlap based approach

5.2 Conceptual Distance based approach

This knowledge based approach is motivated by "Word sense using conceptual density" [1]..Path distance between two synsets is defined as the path length between two synsets in the WordNet graph. An edge on this graph represents hypernymy-hyponymy relations.

i) *Conceptual distance(S1,S2) α Semantic Graph distance(S1,S2)*: Conceptual distance between two synsets is directly proportional to the path length between two synsets.
ii) *Conceptual Closeness(S1,S2) α 1/Path_Distance Of lowest common ancestor of(S1,S2)*: Intuitively conceptual closeness between two synsets is inversely proportional to the height of the lowest common ancestor of two synsets because as the height increases which means the common ancestor becomes more general, the conceptual relatedness between two synsets becomes more vacuous(for example, two synsets related to each other through ENTITY gives no idea about conceptual relatedness between the two synsets.)
iii) From the above discussion we can say that synsets(S1,S2) is closer than synsets(S1,S3) if

{Conceptual Distance(S1,S2) < Conceptual Distance(S1,S3) &&
Conceptual Closeness(S1,S2) > Conceptual closeness(S1,S3)}
|| -----------------(1)
{Conceptual Distance(S1,S2) = = Conceptual Distance(S1,S3) &&
Conceptual Closeness(S1,S2)>Conceptual Closeness(S1,S3)}

```
If (1) holds
S2 is the winner sense
Else
S3 is the winner sense.
```

Ilustrative Example:- The senses of the target word and context words are stored as xml files. The hypernyms of the target word as well as the context words are represented in the form of a graph. Also the hypernyms of the target and context words are stored in xml files. The edges in the graph are actually the hypernymy-hyponymy relations. For example let us consider this sentence:

□□ □□ □□□ □□□□□□□□ □□□□□□□ □□□□□□□ □□□□□ □□□□□ □□□□ □□ □□□□ □□□□□ □□□□ □□ □□□□□□ □□□□(*all rivers issue from mountains and then follow a specific path to an ocean or a large water body*)
Here the target word is □□□□ which has the following senses in WordNet:

Sense1:- □□□□□□, □□□□, □□□□□□, □□□□, □□□□□
Gloss: □□□□ □□□□□□ □□□□ □□□□□ □□□□ □□□□□□□□ □□□□□□□□□□ □□□□□□□□ □□□□ □□□□□ □□□□□□□ □(*a division of an ocean or a large body of salt water partially enclosed by land*)

Example statement:- □□□□□□ □□□□□□□□□ □□□□ □□ / □□□□□ □□□□ □□□□□□ □□□□□□□□ □□□□□□□□ □□□□□□ □□□□□□□ □□□□□ □□□

Sense 2:- □□□□□□, □□□□□, □□□□□□, □□□□, □□□□□□

Gloss:- □□□□ □□□□□□ □□□□□ □□ □□□ □□□□□ □□□□ □□□□ □□□□□□(*an abundant source of information)*
Example statement:- □□□□ □□□□ □□□□□□□ □□□□□□ □□□

Let us disambiguate the polysemous word □□□□ with the monosemous word □□□ as there is only one sense of the word □□□ in Nepali WordNet. The hypernymy-hyponymy graph is as shown below:-

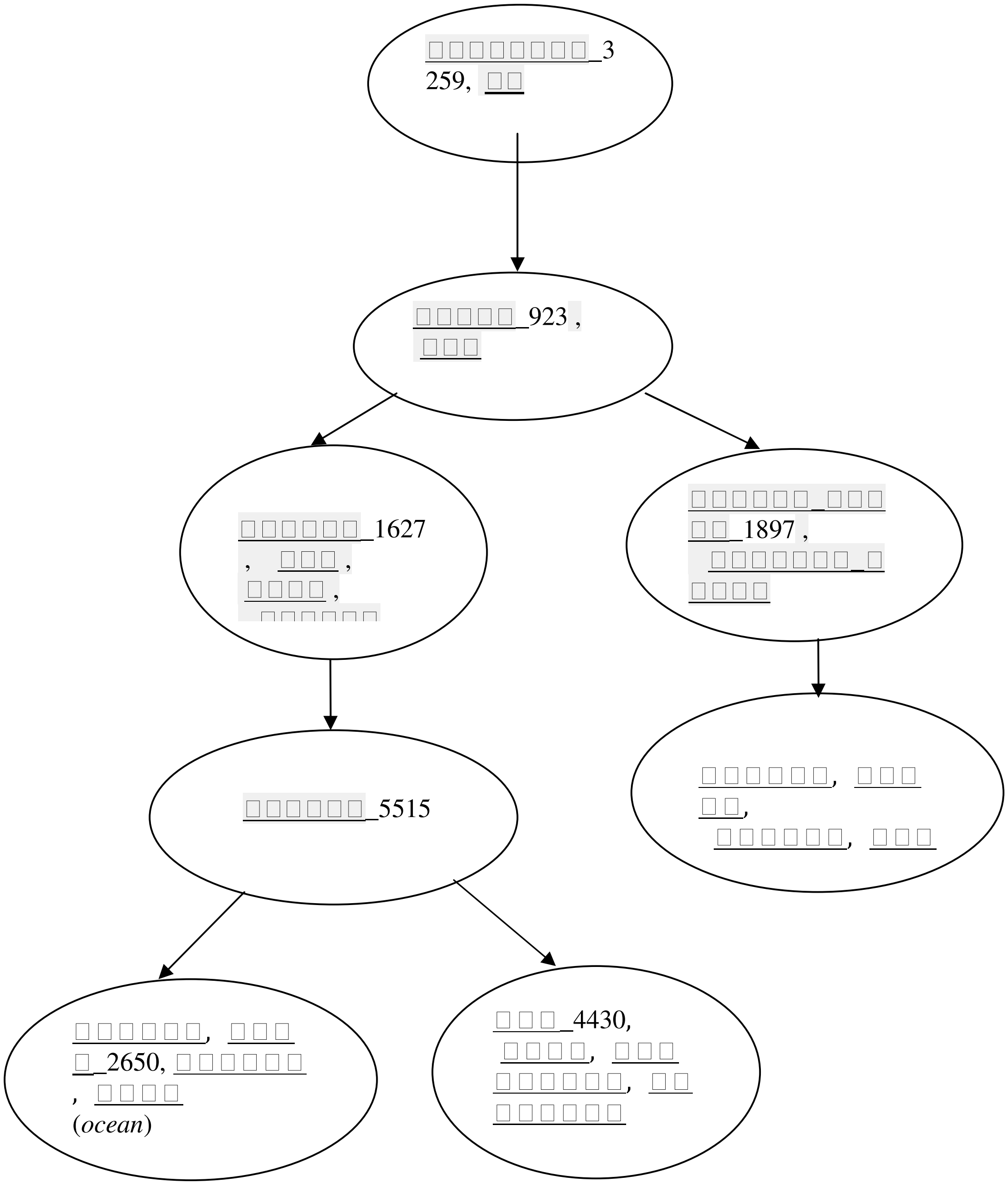


Fig3:Hypernymy-hyponymy graph for the concept सागर and □□□

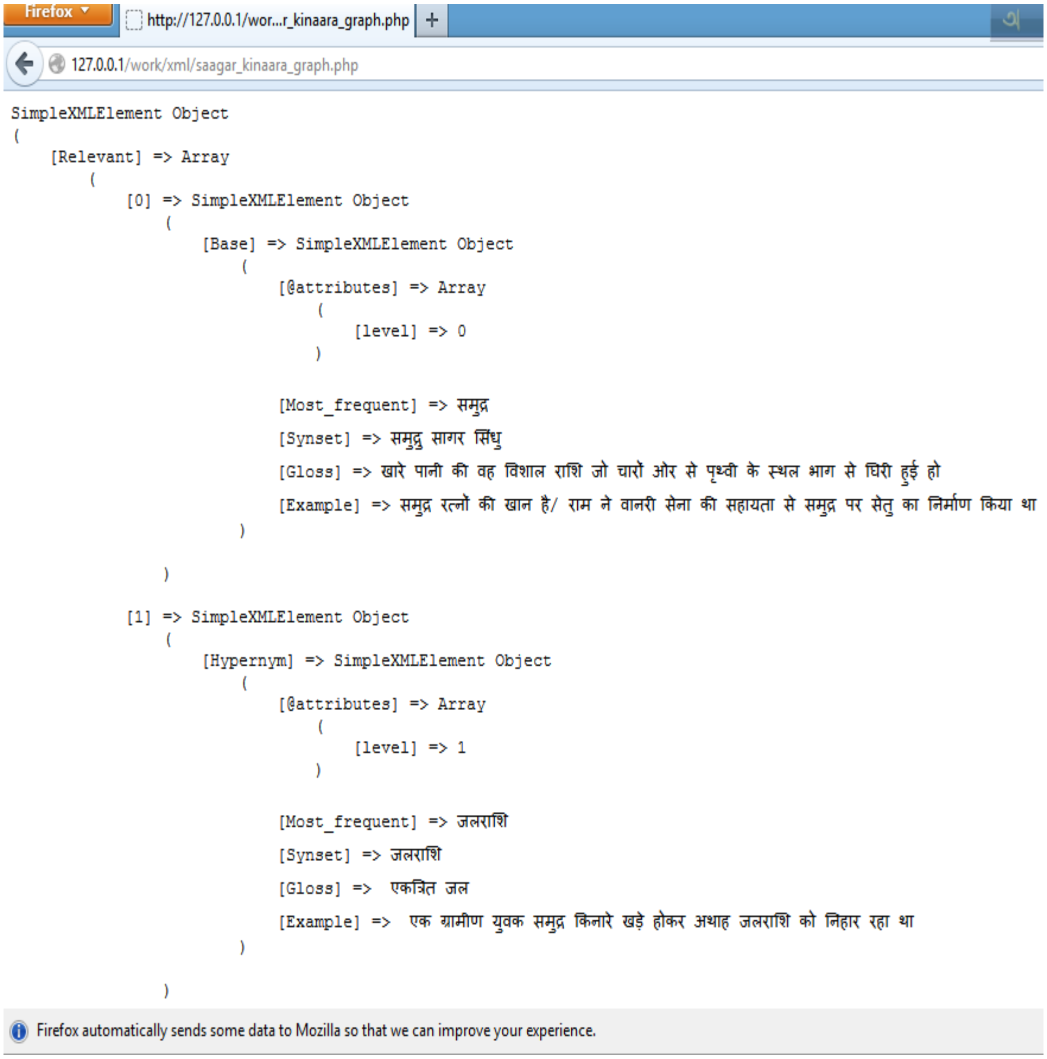
Based on formula given in (1) we can say

a) conceptual distance between (□□□_4430, □□□□_2650) is less than (□□□_4430, □□□□_8231)

b) conceptual closeness between (□□□_4430, □□□□_2650) is more than (□□□_4430, □□□□_8231)

Hence the winner sense is □□□□_2650.(*ocean*)

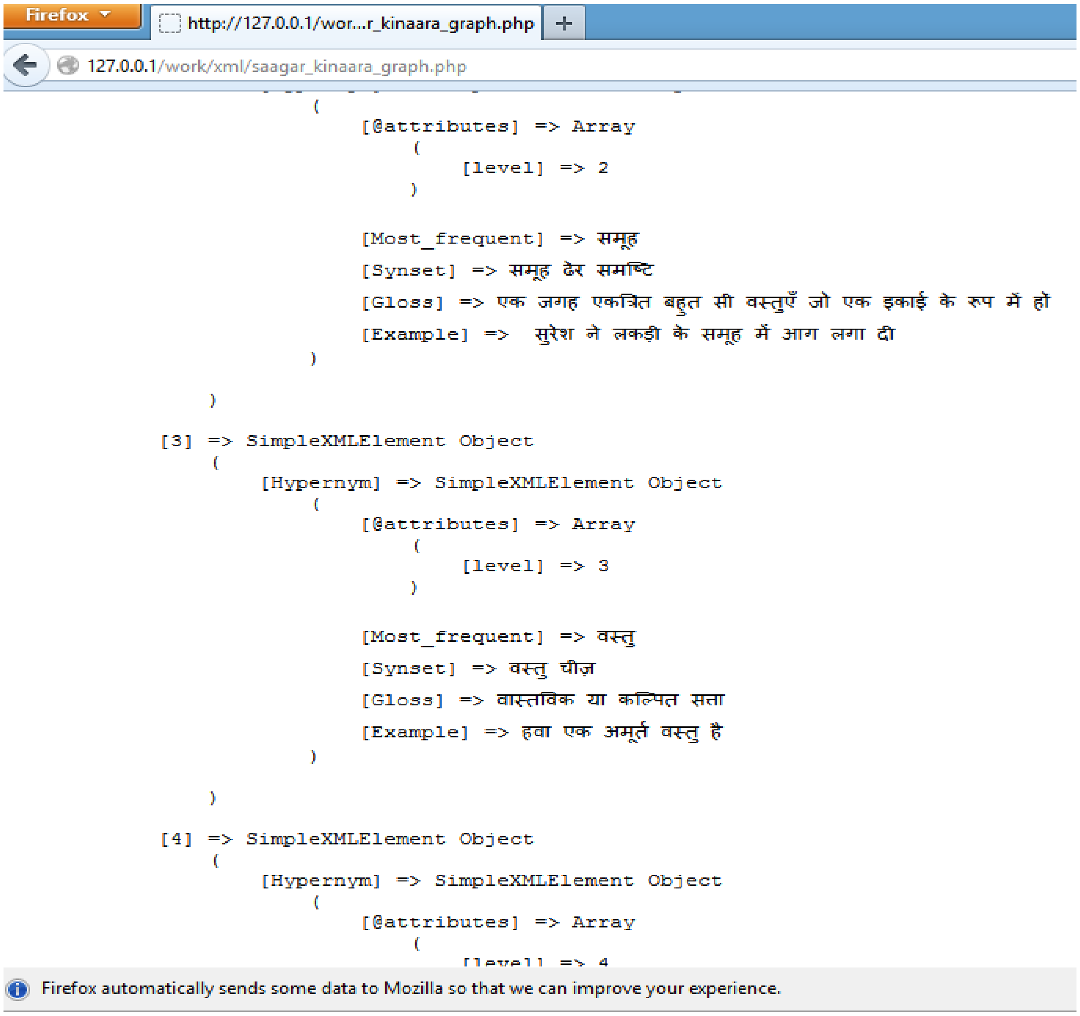
A few screenshots showing the implementation of the graph based approach is shown below


Firefox
http://127.0.0.1/wor...r_kinaara_graph.php
+
127.0.0.1/work/xml/saagar_kinaara_graph.php

```
SimpleXMLElement Object
(
    [Relevant] => Array
        (
            [0] => SimpleXMLElement Object
                (
                    [Base] => SimpleXMLElement Object
                        (
                            [@attributes] => Array
                                (
                                    [level] => 0
                                )

                            [Most_frequent] => समुद्र
                            [Synset] => समुद्र सागर सिंधु
                            [Gloss] => खारे पानी की वह विशाल राशि जो चारों ओर से पृथ्वी के स्थल भाग से घिरी हुई हो
                            [Example] => समुद्र रत्नों की खान है/ राम ने वानरी सेना की सहायता से समुद्र पर सेतु का निर्माण किया था
                        )

                )

            [1] => SimpleXMLElement Object
                (
                    [Hypernym] => SimpleXMLElement Object
                        (
                            [@attributes] => Array
                                (
                                    [level] => 1
                                )

                            [Most_frequent] => जलराशि
                            [Synset] => जलराशि
                            [Gloss] =>  एकत्रित जल
                            [Example] =>  एक ग्रामीण युवक समुद्र किनारे खड़े होकर अथाह जलराशि को निहार रहा था
                        )

                )
```

Firefox automatically sends some data to Mozilla so that we can improve your experience.

Fig4:A screenshot showing example of conceptual distance approach


Firefox
http://127.0.0.1/wor...r_kinaara_graph.php
+
127.0.0.1/work/xml/saagar_kinaara_graph.php

```
            (
                [@attributes] => Array
                    (
                        [level] => 2
                    )

                [Most_frequent] => समूह
                [Synset] => समूह ढेर समष्टि
                [Gloss] => एक जगह एकत्रित बहुत सी वस्तुएँ जो एक इकाई के रूप में हों
                [Example] =>  सुरेश ने लकड़ी के समूह में आग लगा दी
            )

    )

[3] => SimpleXMLElement Object
    (
        [Hypernym] => SimpleXMLElement Object
            (
                [@attributes] => Array
                    (
                        [level] => 3
                    )

                [Most_frequent] => वस्तु
                [Synset] => वस्तु चीज़
                [Gloss] => वास्तविक या कल्पित सता
                [Example] => हवा एक अमूर्त वस्तु है
            )

    )

[4] => SimpleXMLElement Object
    (
        [Hypernym] => SimpleXMLElement Object
            (
                [@attributes] => Array
                    (
                        [level] => 4
```

Firefox automatically sends some data to Mozilla so that we can improve your experience.

Fig5:-Another screenshot showing implementation of conceptual distance and semantic graph approach

5.3 Semantic Graph Distance

Semantic Graph distance is the shortest path between two synsets in the wordnet graph where the edges can be any semantic relation(in case of conceptual distance the edges can only be hypernymy-hyponymy relations).One of the semantic relationships in Nepali Wordnet is (MODIFIES_NOUN). For example this relation holds between two synsets □□□_4003 and □□□□□_744. From the hyponymy relations of □□□□□_744 we arrive at the synset □□□□_6981. If we draw the semantic graph it would look like this:

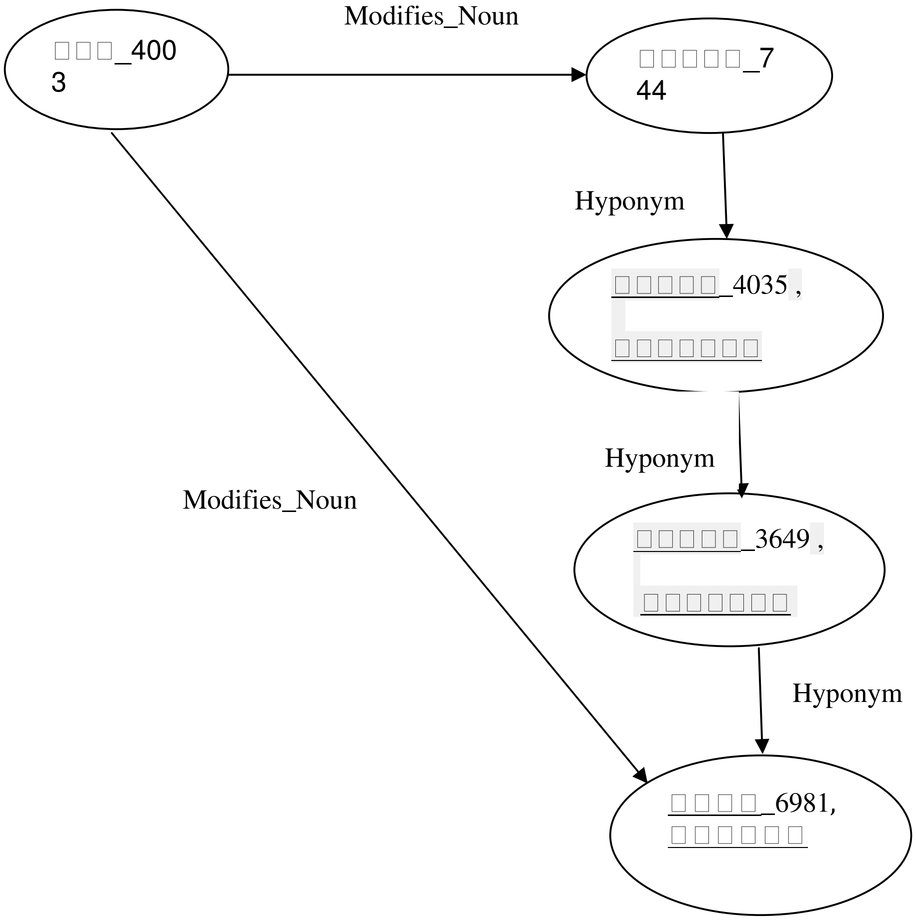


Fig6:-Semantic Graph showing how relation can be inferred between 2 senses when no relationship exists in WordNet.

We can now infer a relation between (□□□_4003) (*categorical,unimpeachable*) and (□□□□_6981, □□□□□□□) (*proof*) which is not present in the WordNet. Intuitively, semantic distance between two synsets is directly proportional to the path length between two synsets.

6.EXPERIMENTS

The system has been tested with a corpus containing Nepali documents obtained from the Indian Language Technology Proliferation and Deployment Center under Ministry of Information Technology, GOI The documents spanned a broad array of topics like novels, short stories, agriculture, health and nutrition etc. Some documents were prepared from online newspaper versions of News of Nepal, Hello Khabar etc. Subjects in the newspapers were classified under Politics, Sports, Health etc. The untagged documents were manually tagged by lexicographers. The total number of words considered were 1663 with 912 nouns and 751 adjectives. The efficiency of the approaches discussed is enumerated in the table below.

TABLE I. COMPARATIVE EFFICIENCY OF NEPALI WSD USING KNOWLEDGE BASED APPROACHES.

	Noun (Correct Sense)	Adjective (Correct Sense)	Accuracy(Noun)	Accuracy (Adjective)
Overlap based approach	482	314	54%(approx)	42%(approx)
Conceptual distance+Semantic Graph distance approach	553	427	62%(approx)	58%(approx)

7.CONCLUSION

From the experimental result it is seen that the performance of overlap based approach is less than the combination of conceptual distance and semantic graph method. It is expected because overlap based approach suffer from sparse overlap. Nevertheless, especially in the case of nouns, the overlap based approach presented here gives better performance than the overlap based approach with machine readable dictionaries because not only the gloss and examples of the target and context synsets are taken but also the gloss and examples from their hypernyms have been taken into consideration. The adjective accuracy is more with the second method as the semantic graph distance score has been taken into account .In future a scoring function to rank the senses based on Hopfield network would be attempted so that performance comparable to state-of the-art could be achieved.

REFERENCES

[1] A. Eneko & G. Rigau, "Word sense using conceptual density" In Proceedings of the 16th International Conference on Computational Linguistics (COLING), Copenhagen, 1996 ndon, vol. A247, pp. 529–551, April 1996.

[2] D. Narayan, D. Chakrabarti, P. Pande and P. Bhattacharyya. An Experience in Building the Indo WordNet - a WordNet for Hindi first International Conference on Global WordNet, Mysore, India, 2002

[3] C. Fellbaum, WordNet: An Electronic Lexical Database. The MIT. 1998

[4] Hindi Wordnet. http://www.cfilt.iitb.ac.in/wordnet/webhwn/

[5] Lee Yoong K., Hwee T. Ng & Tee K. Chia. "Supervised word sense disambiguation with support vector machines and multiple knowledge sources". Proceedings of Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, Barcelona, Spain, 137-140. 2004

[6] L. Dekang. Using syntactic dependency as local context to resolve word sense ambiguity . In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL), Madrid, 64-71.1997

[7] M. Lesk. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone In Proceedings of the 5th annual international conference on Systems documentation, Toronto, Ontario, Canada.1986

[8] R. Navigli, P. Velardi, "Structural Semantic Interconnections: A Knowledge-Based Approach to Word Sense Disambiguation". IEEETransactions On Pattern Analysis and Machine Intelligence. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[9] Véronis Jean. HyperLex: Lexical cartography for information retrieval. Computer Speech & Language, 18(3):223-252, 2004.

[10] D. Walker and R. Amsler The Use of Machine Readable Dictionaries in Sublanguage Analysis. In Analyzing Language in Restricted Domains, Grishman and Kittredge (eds), LEA Press, pp. 69-83, 1986.

[11] Resnik Philip. "Selectional preference and sense disambiguation". In Proceedings of ACL Workshop on Tagging Text with with Lexical Semantics Why, What and How? Washington, U.S.A., 52-57, 1997

[12] M. Rada. "Large vocabulary unsupervised word sense disambiguation with graph-based algorithms for sequence data labeling". In Proceedings of the Joint Human Language Technology and Empirical Methods in Natural Language Processing Conference (HLT/EMNLP), Vancouver, Canada, 411-418, 2005

[13] T. Ng Hwee & Hian B. Lee, "Integrating multiple knowledge sources to disambiguate word senses: An exemplar-based approach". In Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL), Santa Cruz, U.S.A., 40-47,1996.

[14] R. Mohanty, P. Bhattacharyya, P. Pande, S. Kalele, M. Khapra and A. Sharma. Synset Based Multilingual Dictionary: Insights, Applications and Challenges Global Wordnet Conference, Szeged, Hungary, January 22-25, 2008.

[15] Y. David. "Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French". In Proceedings of the 32nd Annual Meeting of the association for Computational Linguistics (ACL), Las Cruces, U.S.A., 88-95. 1994

[16] Y. David. "Unsupervised word sense disambiguation rivaling supervised methods" . In Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL), Cambridge, MA, 189-196.1995

[17] M. Khapra, S. Shah, P. Kedia and P. Bhattacharyya, " Domain-Specific Word Sense Disambiguation Combining Corpus Based and Wordnet Based Parameters", 5th International Conference on Global Wordnet (GWC2010), Mumbai, Jan, 2010.

[18] M. Khapra, P. Bhattacharyya, S. Chauhan, S. Nair and A. Sharma, "Domain Specific Iterative Word Sense Disambiguation in a Multilingual Setting", International Conference on NLP (ICON08), Pune, India, December, 2008

[19] Pedersen,T and S.Banerjee An adapted Lesk algorithm for Word sense disambiguation using WordNet in Proceedings of Third Intrnational Conference on Intelligent Text Processing and Computational Linguistics.Gelbukh.2002

[20] S.Weiss,Learning to Disambiguate.Iinformation Storage and Retrieval,.9:p:33-41.1973

[21] Kelly Edward,F.&J.S.Philip,Computer Recognition of English Word Senses.North Holland,Amsterdam 1975

[22] Piek Vossen,ed,EuroWordNet:A Multilingual Database with Lexical Semantic Networks,Computers and the Humanities 32(2-3),1998

[23] Chakraborty Alok,Roy Arindam,Purkayastha Bipul.,Experiencec in building Nepali WordNet in Proceedings of the 5th Global WordNet Conference,Mumbai,Narosa Publishing House,India,2010