

Tribhuvan University
Institute of Science and Technology

**Graph Based Unsupervised Word Sense Disambiguation with Semantic
Network and Graph Centrality Algorithms**

Dissertation

Submitted to

Central Department of Computer Science and Information Technology

Kirtipur, Kathmandu, Nepal

In partial fulfillment of the requirements
for the Master's Degree in Computer Science and Information Technology

by

Rabin Shrestha

December, 2011

Tribhuvan University
Institute of Science and Technology

**Graph Based Unsupervised Word Sense Disambiguation with Semantic
Network and Graph Centrality Algorithms**

Dissertation

Submitted to

Central Department of Computer Science and Information Technology
Kirtipur, Kathmandu, Nepal

In partial fulfillment of the requirements
for the Master's Degree in Computer Science and Information Technology

by

Rabin Shrestha

December, 2011

Supervisor

Assoc. Prof. Dr. Subarna Shakya

Tribhuvan University
Institute of Science and Technology

Central Department of Computer Science and Information Technology
Kirtipur, Kathmandu, Nepal

Student's Declaration

I hereby declare that I am the only author of this work and that no sources other than the listed here have been used in this work.

.....

Rabin Shrestha

Date: December, 2011

Supervisor's Recommendation

I hereby recommend that this dissertation prepared under my supervision by **Mr. Rabin Shrestha** entitled “**Graph Based Unsupervised Word Sense Disambiguation with Semantic Network and Graph Centrality Algorithms**” in partial fulfillment of the requirements for the degree of M. Sc. in Computer Science and Information Technology be processed for the evaluation.

.....

Assoc. Prof. Dr. Subarna Shakya

Department of Electronics and Computer Engineering,

Institute of Engineering,

Pulchowk, Nepal

Date:

ACKNOWLEDGEMENT

I extend my acknowledgement to my respected teacher and dissertation supervisor Associate Prof. Dr. Subarna Shakya, Department of Electronics and Computer Engineering, Institute of Engineering, Pulchowk, for giving me an opportunity to work under his supervision and for providing me guidance and support throughout this work.

I would like to express my gratitude to the respected teachers Prof. Dr. Shashidar Ram Joshi, Prof. Dr. Onkar P. Sharma, Prof. Sudarshan Karanjeet, Mr. Min Bahadur Khati, Mr. Samujjwal Bhandari, Mr. Bishnu Gautam, Mr. Hemant B. G.C, Mr. Dinesh Bajracharya, and others for granting me broad knowledge within the time period of two years.

Last but not least, I would like to thank my family members and friends for their constant support and encouragement.

Rabin Shrestha

December, 2011

Abstract

This dissertation describes an approach to solve the problem of Word Sense Disambiguation through semantic network of word senses constructed from WordNet as a reference dictionary. Word Sense Disambiguation is the task of assigning a meaning to word, among many available from a dictionary, which is most appropriate depending upon the context of a word. Starting from the single sense vertex graphs, graphs are expanded by adding semantically related senses from WordNet until a graph containing senses of all words in a sentence is constructed. In a graph thus obtained containing senses of all words, graph-based centrality algorithms are applied in order to figure out the most important sense vertex for each word. The method assumes the same word in different sentences may have different meanings. Disambiguation is done by sentence wise, taking sentence as a context for disambiguation. The approach is completely unsupervised and do not require any manually sense tagged resources, but only require a sense repository that provides explicit relation between senses. Weights on edge between senses are assigned using ETF-INF. The method is evaluated by disambiguating texts from standard sense tagged corpora from SemCor and Senseval. The method is suitable for all words domain independent sense disambiguation.

TABLE OF CONTENTS

DETAILS	PAGE NUMBER
CHAPTER 1	
INTRODUCTION	1
1.1 Word Sense Disambiguation	1
1.2 Applications of WSD	2
1.3 Problem Definition	3
1.4 Research Objectives	4
CHAPTER 2	
LITERATURE REVIEW	5
2.1 A Brief History of WSD	5
2.2 Knowledge Sources for WSD	6
2.2.1 Structured Resources	6
2.2.2 Unstructured Resources	7
2.2.3 Collocation Resources	7
2.3 WordNet	7
2.4 Basic Approaches to WSD	10
2.4.1 Dictionary-based or Knowledge-based approaches	10
2.4.2 Corpus-based approaches	10
2.5 Graph Based Approaches to WSD	11
CHAPTER 3	
IMPLEMENTATION	15
3.1 Overview of the phases in Implemented WSD System	15
3.2 Semantic Network Construction	16
3.3 Assigning Edge Weights in Network with ETF-INF	21
3.4 Graph Based Centrality Algorithms	23

3.4.1 Indegree Centrality	23
3.4.2 PageRank	23
3.4.3 HITS	24
CHAPTER 4	
EXPERIMENTS AND EVALUATIONS	26
4.1 Data	26
4.2 The Optimal Size for graphs to expand	27
4.3 Evaluation of the system	28
4.4 Discussion	29
4.5 Limitation of the proposed method	30
CHAPTER 5	
CONCLUSION AND FUTURE STUDY	31
5.1 Conclusion	31
5.2 Future study	31
References	33
Appendix A	[Sample Texts from Test Corpus] 36
Appendix B	[Code for Implementation] 38
Appendix C	[Outputs generated by the Implemented System] 45

LIST OF TABLES

DETAILS	PAGE NUMBER
Table 4.1: Statistics of test data	27
Table 4.2: Results obtained with unweighted semantic network on test data sets	28
Table 4.3: Results obtained with weighted semantic network on test data sets	29
Table 4.4: Pair-wise inter-agreement in test data sets	30

LIST OF FIGURES

DETAILS	PAGE NUMBER
Fig. 2.1: An excerpt of the WordNet semantic network for a sense car_n^1	9
Fig. 2.2: Sample graph built on the set of possible labels for a sequence of four words	12
Fig. 2.3: Graph for the sentence “She drank some milk”	13
Fig. 2.4: Semantic Network construction in the approach of Tsatsoronis, Varlamis, and Norvag	14
Fig. 3.1: Phases in implemented WSD system	15
Fig. 3.2 : Initial phase containing 5 single vertex graphs, which are senses of words w_1, w_2 . S_1 are senses of w_1 , and S_2 are senses of w_2	17
Fig. 3.3 : Expansion phase Round 1, single vertex graphs are expanded	18
Fig. 3.4: Expansion phase Round 2, graphs are on continuous expansion	19
Fig. 3.5: Expansion phase Round 3, a graph with vertices $S_{1.1}, S_{1.2}$ and a graph with vertex $S_{2.1}$ will be merged if $S_{1.2.4.1}$ and $S_{2.1.1.1}$ are same senses	20
Fig. 3.6: A sub graph in the network	22

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
BNC	British National Corpus
DSO	Defence Science Organisation
ETF-INF	Edge Type Frequency Inverse Node Frequency
HITS	Hypertext Induced Topic Search
IE	Information Extraction
IR	Information Retrieval
LDOCE	Longman Dictionary of Contemporary English
MFS	Most Frequent Sense
MT	Machine Translation
NLP	Natural Language Processing
QA	Question Answering
TF-IDF	Term Frequency Inverse Document Frequency
WSD	Word Sense Disambiguation