



Ain Shams University  
Faculty of Computer and Information Sciences  
Computer Science Department

## **Semantic -based Approach for Text Generation**

**Dalia Sayed Fadl**

Computer Science Department,  
Faculty of Computer and Information Sciences,  
Ain Shams University

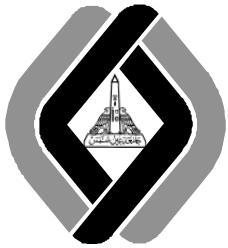
**Under Supervision of**

**Prof. Dr. Mustafa Aref**

Computer Science Department  
Faculty of Computer and Information Sciences  
Ain Shams University

**Dr. Ibrahim F. Moawad**

Information Systems Department  
Faculty of Computer and Information Sciences  
Ain Shams University



Ain Shams University  
Faculty of Computer and Information Sciences  
Computer Science Department

## Semantic -based Approach for Text Generation

A thesis submitted to computer science department. Faculty of computer and information sciences, Ain shams University, in partial fulfillment of the requirements for the degree of Master of Computer and Information Sciences.

By

**Dalia Sayed Fadl**

Approved by the discussion committee:

.....  
**Prof. Dr. Ahmed Sharaf Eldin**

Member

Professor of Computer and Information Sciences  
Faculty of Computer and Information Sciences  
Helwan University  
.....

**Prof. Dr. Mohsen Rashwan**

Member

Professor of comuncation Engineering  
Faculty of Engineering  
Cairo University  
.....

**Prof. Dr. Mostafa Mahmoud Aref**

Member and Supervisor

Professor of Computer Sciences Department  
Faculty of Computer and Information Sciences  
Ain Shams University  
.....

Computer Science Department  
Faculty of Computer and Information Sciences  
Ain Shams University  
Cairo 2012



## انتاج النصوص باستخدام الدلالات اللفظية

داليا سيد محمد  
قسم علوم الحاسب  
كلية الحاسبات و المعلومات  
جامعة عين شمس

بكالوريوس علوم الحاسب  
كلية علوم الحاسب و نظم المعلومات 2007

## تحت اشراف

الاستاذ الدكتور. مصطفى عارف  
قسم علوم الحاسب  
كلية الحاسبات و المعلومات  
جامعة عين شمس

دكتور/ ابراهيم معوض  
قسم نظم المعلومات  
كلية الحاسبات و المعلومات  
جامعة عين شمس

**To my parents, husband, sister and brother.**

## Abstract

*Natural Language Processing enables communication between people and computers. Natural language processing (NLP) is a field of computer science, artificial intelligence and linguistics concerned with the interactions between computers and human (natural) languages. Specifically, it is the process of a computer extracting meaningful information from natural language input and/or producing natural language output. Natural language processing is a very attractive method of human–computer interaction. Natural language understanding is sometimes referred to as an AI-complete problem because it seems to require extensive knowledge about the outside world and the ability to manipulate it. Natural Language Generation (NLG) is sub field of natural language processing which focuses on the generation of written texts in natural language from some underlying semantic representation of information. The objective of natural language generation (NLG) or text generation systems is to produce coherent natural language texts which satisfy a set of one or more communicative goals. To achieve these goals, the generated text should be (among other things). coherent: using well-connected, sensible and comprehensible English; accurate: containing accurate information (or it could lead to the user making false inferences); valid: causing the user to make the desired inferences (for example, telling a naive user that the koala looks like a teddy bear and not telling her that it doesn't behave like one may result in a nasty surprise); understandable: including information which the user can understand; and relevant: including information which is relevant to the current discourse goal and not redundant. In this thesis, a new model to generate an English text from a recent ontology-based semantic representation called (Rich Semantic Graph) is introduced. The developed model can be exploited in Text Summarization, Machine Translation and Information Retrieval applications. Because of generating multiple texts, the phase of text evaluation is developed in our model to evaluate the final multiple texts based on the most frequently used words using WordNet ontology and the relations between sentences. In this model, WordNet ontology is used to generate multiple texts according to the word synonyms. Also, the model enables users to determine the output text style by selecting one of two writing styles (Cause-Effect and Description-Narration). Finally, the model evaluates the generated texts to rank them based on two criteria: most frequently used words and discourse sentence relations.*

## Acknowledgments

Thanks to **Allah** before and after, and heartily thankful to my supervisors; **Professor Mostafa Aref** for his encouragement guidance and support from the initial to final level, enabled me to develop and understand Natural language Generation. He was not just a supervisor but a father in all the situations. And I would like to show my gratitude to my great supervisor; **Dr Ibrahim Fathy** without his patient guidance and support I would not make this work. I would like to thank my family for all their love and encouragement; My parents who raised me with a love of science and supported me in all my pursuits; My brother Mohamed and my sister Dina. And, this thesis would not have been possible without my loving husband Tarek Khattab, who helped, supported, encouraged and believed in me all the time at all the situations. Thank you is very small word to express my feeling. I am thankful to my friends and my colleges how supported me. I would like to appreciate the help offered by my great colleague Mohamed mamdoh how helped me so much every time I was in need for any help.

## **Publications**

Parts of this thesis are published as original papers in the following references:

1. Mostafa Aref, Ibrahim Fathy, Dalia Sayed “Natural language Generation Sentence Planner” International Conference on intelligent computing and information systems ICICIS 2011 p22, Faculty of computer and information science Ain shams University, Cairo, Egypt.
2. Dalia Sayed, Mostafa Aref, Ibrahim Fathy “Text Generation Model from Rich Semantic Representations” Egypt Society of Language Engineering ESOLEC’ 2011 Pages 68-67, Faculty of engineering Ain shams university, Cairo, Egypt.
3. Ibrahim Moawad, Dalia Sayed, Mostafa Aref “Rich Semantic Representation Based Approach for Text Generation” INFOS 2012.

# TABLE OF CONTENTS

|                                                                 | Page      |
|-----------------------------------------------------------------|-----------|
| <b>ABSTRACT.....</b>                                            | <b>1</b>  |
| <b>ACKNOWLEDGEMENTS.....</b>                                    | <b>2</b>  |
| <b>TABLE OF CONTENTS.....</b>                                   | <b>3</b>  |
| <b>LIST OF FIGURES.....</b>                                     | <b>4</b>  |
| <b>LIST OF TABLES.....</b>                                      | <b>5</b>  |
| <b>LIST OF ABBREVIATIONS.....</b>                               | <b>6</b>  |
| <b>CHAPTER 1 INTRODUCTION.....</b>                              | <b>8</b>  |
| 1.1 Overview.....                                               | 9         |
| 1.2 Aims and Contributions.....                                 | 11        |
| 1.3 Structure of the thesis.....                                | 11        |
| <b>CHAPTER 2 BACKGROUND AND RELATED WORK.....</b>               | <b>13</b> |
| 2.1 Text generating process steps.....                          | 14        |
| 2.2 Problems in generating NLG systems .....                    | 16        |
| 2.3 Semantic text representation .....                          | 17        |
| 2.4 Ontology.....                                               | 17        |
| 2.4.1 Ontology in text generation.....                          | 18        |
| 2.4.2 The Relation between Ontologies and Natural Language..... | 18        |
| 2.5 Generated text Evaluation.....                              | 19        |
| 2.6 Text Generation approaches.....                             | 20        |



|                                                  |                                                     |           |
|--------------------------------------------------|-----------------------------------------------------|-----------|
| 2.6.1                                            | Traditional Generation approaches.....              | 20        |
| 2.6.2                                            | Ontology based approach.....                        | 21        |
| 2.7                                              | Related Work.....                                   | 22        |
| <b>CHAPTER 3 DESIGN.....</b>                     |                                                     | <b>26</b> |
| 3.1                                              | NLG Model Conceptual View.....                      | 27        |
| 3.2                                              | THE NLG System Model Overview.....                  | 28        |
| 3.2.1                                            | Text Planning .....                                 | 30        |
| 3.2.2                                            | Sentence Planning.....                              | 30        |
| 3.2.3                                            | Surface Realization.....                            | 33        |
| 3.2.4                                            | Writing Styles Selected Essay Generation phase..... | 34        |
| 3.2.5                                            | Evaluation.....                                     | 34        |
| 3.3                                              | Case Studies.....                                   | 36        |
| <b>CHAPTER 4 NLG Model Implementation.....</b>   |                                                     | <b>48</b> |
| 4.1                                              | Output.....                                         | 49        |
| 4.2                                              | Analysis.....                                       | 64        |
| <b>CHAPTER 5 Conclusion and Future Work.....</b> |                                                     | <b>66</b> |
| 5.1                                              | Conclusion.....                                     | 67        |
| 5.2                                              | Contributions.....                                  | 67        |
| 5.2                                              | Future Work.....                                    | 68        |

## LIST OF ABRIVIATIONS

| Abbreviation |                                                            | Page |
|--------------|------------------------------------------------------------|------|
| <b>W</b>     | Weight.....                                                | 30   |
| <b>E</b>     | Existence of Synonyms in rich semantic graph.....          | 30   |
| <b>NR</b>    | Synonym WordNet Rank.....                                  | 30   |
| <b>TR</b>    | Total Values of All Synonym Ranks.....                     | 30   |
| <b>NGS</b>   | WordNet Group By Similarity for Synonyms.....              | 30   |
| <b>TG</b>    | Total Number of Groups by Similarity for all Synonyms..... | 30   |
| <b>DS</b>    | Discourse Relations.....                                   | 35   |
| <b>WNR</b>   | WordNet Rank.....                                          | 35   |
| <b>OWL</b>   | Ontology Web language.....                                 | 31   |
| <b>RDF</b>   | Resource Description Framework.....                        | 31   |
| <b>RSG</b>   | Rich Semantic Graph.....                                   | 35   |

## LIST OF TABLES

| Table                      | Page |
|----------------------------|------|
| 4.1 Case 1Outputs.....     | 50   |
| 4.2 Case 2 Outputs .....   | 52   |
| 4.3 Case 3 Outputs.....    | 54   |
| 4.4 Case 4 Outputs.....    | 56   |
| 4.5 Case 5 Outputs.....    | 58   |
| 5.6 Case 6 Outputs.....    | 60   |
| 4.7 Case 7 Outputs.....    | 62   |
| 4.8 NLG Model Outputs..... | 64   |

## LIST OF FIGURES

| Figure                                                                  | Page |
|-------------------------------------------------------------------------|------|
| 3.1 NLG Model Conceptual View .....                                     | 28   |
| 3.2 Natural Language Generation Model .....                             | 29   |
| 3.3 Discourse Structuring Algorithm .....                               | 31   |
| 3.4 Example of Subject Grouping.....                                    | 31   |
| 3.5 Example of Predicate Grouping.....                                  | 31   |
| 3.6 Discourse Structuring Relations Example.....                        | 32   |
| 3.7 Discourse Structuring Relations Algorithm.....                      | 33   |
| 3.8 Referring Expression Generation Algorithm.....                      | 33   |
| 3.9 Evaluation Algorithm.....                                           | 35   |
| 3.10 Model Input (Student Rich Semantic Graph) .....                    | 36   |
| 3.11 Model Output1 (two Description-Narration Styled Paragraphs).....   | 36   |
| 3.12 Model Output2 (One Cause-Effect Styled Paragraph).....             | 36   |
| 3.13 Lexicalization Output (Sample of Selected Object Synonyms).....    | 37   |
| 3.14 Discourse Structuring Output (Pseudo-Sentences Sample).....        | 37   |
| 3.15 Aggregation Output (Sample of Semi-Paragraphs).....                | 38   |
| 3.16 Referring Expression Output (Enhanced Semi-Paragraphs sample)..... | 38   |
| 3.17 Surface Realization Output (Sample of Paragraphs).....             | 39   |
| 3.18 Model Input (Family Rich Semantic Graph).....                      | 39   |
| 3.19 Model Output1 (two Description-Narration Styled Paragraphs).....   | 40   |
| 3.20 Model Output2 (One Cause-Effect Style Paragraph).....              | 40   |
| 3.21 Lexicalization Output (Sample of Selected Object Synonyms).....    | 40   |

|      |                                                                    |    |
|------|--------------------------------------------------------------------|----|
| 3.22 | Discourse Structuring Output (Pseudo-Sentences Sample).....        | 41 |
| 3.23 | Aggregation Output (Sample of Semi-Paragraphs).....                | 41 |
| 3.24 | Referring Expression Output (Enhanced Semi-Paragraphs sample)..... | 42 |
| 3.25 | Surface Realization Output (Sample of Paragraphs).....             | 42 |
| 3.26 | Model Input (Travel Rich Semantic Graph).....                      | 43 |
| 3.27 | Model Output1 (two Description-Narration Styled Paragraphs).....   | 43 |
| 3.28 | Model Output2 (One Cause-Effect Style Paragraph).....              | 43 |
| 3.29 | Lexicalization Output (Sample of Selected Object Synonyms).....    | 44 |
| 3.30 | Discourse Structuring Output (Pseudo-Sentences Sample).....        | 44 |
| 3.31 | Aggregation Output (Sample of Semi-Paragraphs).....                | 45 |
| 3.32 | Referring Expression Output (Enhanced Semi-Paragraphs sample)..... | 45 |
| 3.33 | Surface Realization Output (Sample of Paragraphs).....             | 46 |
| 4.1  | Case 1 Input.....                                                  | 49 |
| 4.2  | Case 1 Output.....                                                 | 49 |
| 4.3  | Case 1(Generated Sentences).....                                   | 50 |
| 4.4  | Case 1 (Generated Paragraphs).....                                 | 50 |
| 4.5  | Case 2 Input.....                                                  | 51 |
| 4.6  | Case 2 Output.....                                                 | 51 |
| 4.7  | Case 2 (Generated Sentences).....                                  | 52 |
| 4.8  | Case 2(Generated Paragraphs).....                                  | 52 |
| 4.9  | Case 3 Input.....                                                  | 53 |
| 4.10 | Case 3 Output.....                                                 | 53 |
| 4.11 | Case 3(Generated Sentences).....                                   | 54 |

|      |                                    |    |
|------|------------------------------------|----|
| 4.12 | Case 3(Generated Paragraphs).....  | 54 |
| 4.13 | Case 4 Input.....                  | 55 |
| 4.14 | Case 4 Output.....                 | 55 |
| 4.15 | Case 4 (Generated Sentences).....  | 56 |
| 4.16 | Case 4 (Generated Paragraphs)..... | 56 |
| 4.17 | Case 5 Input.....                  | 57 |
| 4.18 | Case 5 Output.....                 | 57 |
| 4.19 | Case 5(Generated Sentences).....   | 58 |
| 4.20 | Case 5(Generated Paragraphs).....  | 58 |
| 4.21 | Case 6 Input.....                  | 59 |
| 4.22 | Case 6 Output.....                 | 59 |
| 4.23 | Case 6 (Generated Sentences).....  | 60 |
| 4.24 | Case 6 (Generated Paragraphs)..... | 60 |
| 4.25 | Case 7 Input.....                  | 61 |
| 4.26 | Case 7 Output.....                 | 61 |
| 4.27 | Case 7(Generated Sentences).....   | 62 |
| 4.28 | Case 7(Generated Paragraphs).....  | 62 |
| 4.29 | Cases from 1 to 4.....             | 64 |
| 4.30 | Cases from 5 to 7.....             | 64 |

# 1

# Introduction

## **CHAPTER OUTLINE**

---

- **Overview**
- **Aims and Contributions**
- **Structure of the thesis**