قال عزّ و جلّ:

" وَاتَّـقُواْ اللّه وَيُعَلِّمُكُـمُ اللّهُ وَاللّهُ بِكُـلِّ شَيءٍ عَلِيـم "

سورة البقرة ، الآية 282

To the most precious people ALLAH had blessed me with.

I wish if I could have back to be share this special moment with me.

My father Farah My mother Faiza thanks you for everything you have done to make me who I am.

My sister Ayaa

My ALLAH bless you and have mercy on your souls.

Acknowledgement

First and foremost, I must thank ALLAH (S.W.T). For His unlimited bounties and innumerous graces for helping me in all life.

I would like to gratefully and sincerely thank my supervisor Dr. Albaraa AbuObieda for his guidance, understanding, patience, and most importantly, his friendship during my graduate studies. I am grateful to him for giving his time and his guidance, and for helping me overcome the challenges I faced in my study. For everything you've done for me, Dr. Albaraa, I thank you.

I am grateful to all my graduate and under graduate colleagues, friends, staff, and lecturers in Faculty of Computer Science & information Technology at Sudan University of Science & Technology and International University of Africa.

I wish to extend my grateful appreciation to all my friends and all those who have contributed directly and indirectly to the preparation of this study, In particular to my aunt Fatima Abdo, Dr. Ahmed Hamza, Hadiya AbdoAlgadir, Dr. Arafat AbdoAlgadir, Huda khidir, Hala Hamad and Bakry Alsafi.

I would like to express my deep gratitude and thanks to my family for their patience, support and prayers.

Abstract

Sentence compression is the task of compressing a long sentence into a short one. It retains the important contents and in the meantime it generates grammatical short sentences. This research proposed an improved semantic-based method to investigate semantic role labels in sentence compression. Moreover, in the process of compression, the approach is applied in Ziff–Davis dataset. It includes short sentences as summary for a corresponding long ones. In the process of compression we proposed to design what so called semantic patterns, so each sentence has been compressed in three ways. To evaluate the result, we used Jaccard similarity to compare between our approach and human short sentences on the Ziff–Davis dataset. The results showed that our proposed semantic-based method is recommended to be used for sentence compression.

الخلاصة

تطبيق معالجة اللغات المسمى ب" ضغط الجمل" يهدف إلى تقليص الجمل ذات العدد الكبير من الكلمات وتلخيصها إلى جمل قصيرة. هذه التطبيقات يجب ان تحتفظ بالأجزاء الرئيسية للجملة والتي تحمل خبر الجملة, وفي نفس الوقت تقوم بتوليد أو انشاء جمل قصيرة ذو طابع نحوي سليم. في هذا البحث تم استخدام عينه بيانات مخصصه لهذا الغرض وتسمى ب"Ziff Davis". وهذه البيانات المعدة لإجراء الإختبار تحتوي على جمل طويلة تليها جمل قصيره معدة من قبل خبير باللغة.

في عملية تقليص الجمل صصمنا مايسمى ب "patterns" ولذلك عملية تقليص الجمل يتم بثلاث طرق. لقياس درجة كفاءة الطريقة المقترحة من قبلنا, لقد قمنا بإستخدام القياس المسمى ب"Jaccard" لحساب درجة التشابه بين الجمل القصيرة المولده عن طريق طريقتنا المقترح وعن طريق الجمل المصاحب في عينة بيانات "Ziff Davis". النتائج النهائية لهذا البحث أوضحت أن طريقتنا المقترحة والتي تستخدم الطرق الدلالية للمعاني يمكن إعتماداها كآلية لضغط الجمل وتقليصها.

TABLE OF CONTENT

TITLE	P.NO
الآيه……	I
DEDICATION……	II
ACKNOWLEDGEMENT ……	III
ABSTRACT…….....	Ii
الخلاصة……....	Iii
TABLE OF CONTENTS ……	Vi
LIST OF FIGURE……..	ix
LIST OF TABLE ……...	x
LIST OF ABBREVIATIONS……	xi

CHAPTER1

1INTRODUCTION ..... ……..……1

1.1 BACKGROUND ..... ……..……1

1.2 PROBLEM STATEMENT ..... ……..……1

1.3 RESEARCH OBJECTIVES ..... ……..……2

1.4 RESEARCH QUESTION ..... ……..……2

1.5 RESEARCH HYPOTHESES ..... ……..……3

1.6 RESEARCH SCOPE ..... ……..……3

1. 7 RESEARCH SAGNIFICANCE ..... ……..4

CHAPTER2

2 LITERATURE REVIEW & RELATED APPLICATION

2.1 INTRODUCTION ..... ……..…...5

2.2 INTRODUCTION TO NATURAL LANGUAGE PROCESSING……...6

2.3 INFORMATION RETRIEVAL ..... ……....7
2.4 TEXT SUMMARIZATION (TS)…….8
2-4-1 TYPES OF TS ..... ……..8
2.5 SENTENCE COMPRESSION……...9
2.6 Works USED NOISY CHANNEL FOR SENTENCE COMPRESSION…….....10
2.7 INTRODUCTION TO SEMANTIC ROLES LABELING (SRL)……12
2.7.1 SEMANTIC ROLES……...13
2.7.2 IMPORTANT OF SRL……...14
2.7.3 PROPBANK PROJECT……..…14
2.7.4 ROLE EXTRACTOR TOOLS……...15
2.7.5 WORKS USED SRL TO COMPRESS SENTENCES…….16
2.8 DATA SETES…….17
2.9 SENTENCES SIMILIRITY MEASURES…….17
2.9.1 JACCARD SIMILARITY COEFFICIENT……....18
2.10 SUMMARY……19

CHAPTER3

3 METHODOLOGY

3.1 INTRODUCTION……20
3.2 METHODOLOGY……..……20

CHAPTER4

4 **DESIGNING**
4.1INTRODUCTION..... ……24
4.1.1 PHASE 1: ROLE EXTRACTION……....26
4.1.2 PHASE 2: CALCULATE THE WEIGHTS……27
4.1.3 PHASE 3: SENTENCE COMPRESSION……..……..28

CHAPTER5

5 RESULTS AND DISCUSSION

5.1Description……30
5.2 CONCLUSIONS……...44

REFERENCES……..45

LIST OF FIGURES

FIGURE NO	TITLE	PAGE
2.1	The role extractor for sentence	15
3.1	The general framework of our proposed method	21
3.2	All steps of methodology to compression sentences.	22
4.1	Role extraction for sentence one in excel file	24
4.2	Role extraction for sentence tow in excel file	25
4.3	weight of roles for original sentences	26
4.4	Weight of roles for short sentences	26
4.5	Representation the role extractor for one sentence	31
4.6	Compression sentences depend on main verb.	32

LIST OF TABLES

TABLE NO	TITLE	PAGE
2.1	Argument modifiers in PropBank	13
4.1	meaning of some SRLs	25
4.2	The obtained weight for all roles extracted from original sentences.	27
4.3	The obtained weights for all roles extracted from the human short sentences.	28
4.4	The contents of Patterns	30
5.1	Similarity scores of all reduced Sentences using Pattern 1, Pattern 2 and Pattern 3 against Human-Short Sentences.	34
5.2	Numbers of rate increment in pattern than others	37
5.3	average similarity for patterns and Similarity Rate’s when it equals one	38
5.4	compare the performance result of both Pattern3 and Human against Original Sentences	39
5.5	illustrates the evaluate the result of our system	41

LIST OF ABBREVIATIONS

NLP	-	Natural Language Processing
MT	-	Machine Translation
IR	-	Information Retrieval
TS	-	Text Summarization
SRL	-	Semantic Roles Labeling
SDS	-	Single- Document Summarization
MDS	-	Multi-Document Summarization