

**Deanship of Graduate Studies
Al-Quds University**



**Extracting Learning Objectives from Arabic Text Based
on Bloom's Taxonomy**

Tareq Issam Abdel-Azeez Ali

M.Sc. Thesis

Jerusalem-Palestine

1440-2019

**Extracting Learning Objectives from Arabic Text Based
on Bloom's Taxonomy**

Prepared by:

Tareq Issam Abdel-Azeez Ali

B.Sc. Computer Science / Najah university /Palestine

Supervisor: Dr. Badie Sartawi

**A thesis submitted in fulfillment of requirements for the
degree of Master of Computer Science of Al-Quds
University.**

1440/2019

Al-Quds University
Deanship of Graduate Studies
Computer Science



Thesis Approval

Extracting Learning Objectives from Arabic Text Based on Bloom's Taxonomy

Student Name: Tareq Issam Abdel-Azeez Ali

Registration No: 21411042

Supervisor: Dr. Badie Sartawi

Master thesis submitted and accepted date: 23/6 /2019

The names and signatures of the examining committee members are as follows:

1-Head of committee: Dr. Badie Sartawi

signature: 

2- Internal Examiner: Dr. Rashid Jayousi

signature: 

3- External Examiner: Dr. Radwan Tahboub

signature: 

Jerusalem- Palestine

1440/2019

Dedication

This thesis is dedicated to my parents, my brothers and sisters and my wife who have always loved me unconditionally and who was good examples have taught me to work hard for the things that i aspire to achieve I am truly grateful for having in my life. This work is also dedicated to my teachers who have always been a constant source of support and encouragement during the challenges of my whole college life.

Declaration

I certify that this thesis submitted for the degree of Master of Science is the result of my own research, except where otherwise acknowledged, and that this thesis (or any part of the same) has not been submitted for a higher degree to any other university or institution.

Signed

Tareq Issam Abdel-Azeez Ali

Date: 23 / 6 / 2019

Acknowledgment

First of all, I have to thank Allah and praise him so much for his infinite graces, Allah gave me patience and assistance to accomplish this research, and Allah taught me what I did not know, so Alhamdulillah.

Many thanks and sincere gratefulness to the teachers Dr. BADIE SARTAWI for his efforts, guidance and assistance during the work in this thesis and before, and I would like to thank the rest of the thesis committee and the department of computer science for all their support during my research.

Special thanks also to my parents, brothers, sisters, my wife and my lovely little daughter 'Reem' for their continued love, trust and support, without them I would never have been able to achieve what I have achieved.

Abstract

Arabic language one of the most complex and comprehensive languages, and the need to find a uniform and accurate pattern for analysis and understanding the language has become an urgent necessity in light of this technological acceleration, which can help in several areas, including education and research, and this may be helpful in trying to automating the education using the methodology of learning objectives and learning outcomes. The aim of this thesis is to analyze Arabic text in order to extract the meaningful learning objectives from the text based on the first two level from bloom's Taxonomy, remembering and understanding, using two main methodologies the first one by building Arabic Ontology (a primitive tree in this stage), which include Arabic words classification a fixed and unified classification based on the origin of the word, the second methodology is by adoption a standard templates for the learning objectives each template include several keywords and indicators then using three stages. First stage perform pre-processing operations on the text such as divide the text into (paragraphs, sentences, words), remove/set diacritic marks, exclude some words before processing, the second stage finding the indicators for each learning objective template and matching the sentences or paragraph with this template taking into account that there is a degree of matching depend on the number of template indicators found, the number of nouns in the sentence, and some symbols and numbers, the final stage is to extract all the words from the text which fall in the same classification in the Arab Ontology tree, this classification can also be used to exclude certain words to improve the degree of matching. The methods which used to test and evaluate the accuracy of approach using calculate precision, recall and f-measure. The results show satisfactory results for some learning objectives like extracting definitions with precision = 95% and recall = 90%, while the results for other learning objectives were fairly results.

Keywords – Learning Objectives, Arabic Ontology, Bloom's Taxonomy

إستخلاص الأهداف التعليمية من النص العربي على أساس تصنيفات بلوم

إعداد: طارق عصام عبد العزيز علي

إشراف: د. بديع سرطاوي

الملخص

تعد اللغة العربية واحدة من أكثر اللغات تعقيداً وشمولية والحاجة إلى إيجاد نمط موحد ودقيق للتحليل وفهم اللغة أصبحت ضرورة ملحة في ضوء هذا التسارع التكنولوجي ، الذي يمكن أن يساعد في العديد من المجالات ، بما في ذلك التعليم والبحث ، وقد يكون ذلك مفيداً في محاولة أتمتة التعليم باستخدام منهجية أهداف التعلم ونتائج التعلم.

الهدف من هذه الرسالة هو تحليل النص العربي من أجل استخلاص الأهداف التعليمية الهادفة من النص على أساس أول مستويين من تصنيف بلوم، التذكر والفهم باستخدام منهجين رئيسيين أولهما عن طريق بناء علم الوجود العربي (شجرة بدائية) في هذه المرحلة، والتي تتضمن تصنيف الكلمات العربية تصنيفاً ثابتاً وموحداً يستند إلى أصل الكلمة، والمنهجية الثانية هي اعتماد نماذج قياسية لأهداف التعلم تشمل كل نموذج عدة كلمات رئيسية ومؤشرات ثم استخدام ثلاث مراحل. المرحلة الأولى تقوم بعمليات المعالجة المسبقة في النص مثل تقسيم النص إلى (فقرات، جمل، كلمات)، إزالة / تعيين علامات التشكيل ، استبعاد بعض الكلمات قبل البحث، المرحلة الثانية إيجاد المؤشرات لكل قالب من قوالب الاهداف التعليمية ومطابقة الجمل أو الفقرة مع هذا القالب مع الأخذ بعين الاعتبار أن هناك درجة من المطابقة تعتمد على عدد مؤشرات القوالب التي تم العثور عليها، وعدد الأسماء في الجملة، وبعض الرموز والأرقام، المرحلة النهائية هي استخراج كل الكلمات من النص الذي تقع في نفس التصنيف في شجرة الوجود العربي، يمكن استخدام هذا التصنيف أيضاً لاستبعاد كلمات معينة لتحسين درجة المطابقة.

الأساليب التي تستخدم لاختبار وتقييم الطريقة باستخدام حساب دقة النتائج حيث تظهر نتائج مرضية لبعض الأهداف التعليمية مثل استخراج التعريفات بدقة = 95% ونسبة إرجاع = 90% ، بينما كانت نتائج أهداف التعلم الأخرى نتائج معقولة..

الكلمات الأساسية - الأهداف التعليمية ، شجرة الأنطولوجيا العربية ، تصنيفات بلوم

Table of Contents

Declaration.....	I
Acknowledgment.....	II
Abstract.....	III
الملخص.....	IV
Table of Contents.....	V
List of Tables.....	VIII
List of Figures.....	IX
List of Abbreviations.....	XI
Chapter 1.....	2
Introduction.....	2
1.1 Introduction.....	2
1.2 Statement of the Problem.....	3
1.3 Research question.....	4
1.4 Contribution.....	4
1.5 Scope and Limitations.....	5
1.6 Research Methodology.....	5
1.7 Research Outline.....	8
Chapter 2.....	10
Related Work.....	10
2.1 Literature Review.....	10
2.2 Theoretical Concern.....	12
Chapter 3.....	15
Theoretical and Technical Foundation.....	15
3.1 Ontology.....	15
3.1.1 Structure of Arabic Ontology.....	17
3.1.2 Contribution of ontology tree in the extraction of learning objectives.....	17

3.2 Global Terminology Recognition	18
3.3 Extracting Learning objectives based on Blooms Taxonomy	19
3.3.2 Knowledge Level Sections	22
3.4 Constructing Templates for Learning Objectives	23
3.5 Arabic Text Processing	24
3.5.2 Diacritical Marks	25
3.5.3 Original Word Extract.....	25
3.6 Performance Evaluation.....	26
3.7 Summary	27
Learning objectives Extraction Model.....	28
Chapter 4.....	29
Learning objectivesExtraction Model.....	29
4.1 The Basic Model.....	29
4.2 Constructing Ontology Tree	32
4.3 Collecting Data and Pre-processing.....	33
4.3.1 Collecting and classifying Data	33
4.3.2 Pre-processing.....	35
4.4 Global Terminologies	37
4.5 Templates for Learning Objectives.....	38
4.5.1 Remembering level	38
4.5.2 Understanding level	40
4.6 Summary	42
Chapter 5.....	44
Implementation	44
5.1 Programs and Tools	44
5.2 Implement Pre-processing Stage.....	45
5.2.1 Splitting.....	45

5.2.2 Normalization	48
5.2.3 Tokenization	50
5.3 Ontology Based Extraction	50
5.4 Extracting Global Terminologies.....	52
5.5 Implement the General Templates to Extract LO	53
5.5.1 Remembering level	53
5.5.2 Understanding level	57
5.6 Summary	60
Chapter 6.....	62
Experimental Results and Evaluation	62
6.1 Experiments Setup	62
6.2 Preparation	63
6.3 Arabic Text Files and Scenarios	65
6.3.1 Sience Content.	65
6.3.2 Text content of history.	66
6.3.3 Text content of general topics.....	67
6.4 Evaluation of Results	68
6.5 Discussion	73
6.6 Summary	74
Chapter 7.....	76
Conclusions.....	76
7.1 Summary	76
7.2 Future Work	77
References.....	78
Appendix: code to extract learning objectives.....	82

List of Tables

Table 3.1: Global Terminology Classification.....	18
Table 3.2: Learning objectives for each level.....	21
Table (4.1): Pronoun Classifications.....	34
Table (4.2): Huruf Classifications.....	34
Table (4.3): Splitting Indicators.....	36
Table (6.1): Computer Specifications	63
Table (6.2): Summary of Precision, Rrecall and F-measure for All Text Contents	73

List of Figures

Figure 1.1: The Approach to Extract Learning Objectives.....	8
Figure 3.1: A Section of Ontology Tree – in English.....	16
Figure 3.2: A Section of Ontology Tree – in Arabic.....	21
Figure 3.3: Andersons Taxonomy.....	21
Figure 3.4: Model Flowchart for Arabic Text Processing.....	24
Figure (4.1): Model Flowchart for Extracting LO.....	32
Figure (4.2): Constructing Ontology Tree.....	33
Figure (4.3): Pre-processing Stage.....	37
Figure (4.4): An example of Tokenization Process.....	37
Figure (4.5): General Template for Definitions.....	38
Figure (4.6): General Template for Enumerations.....	39
Figure (4.7): General Template for Comparisons.....	41
Figure (5.1): Original Arabic text before splitting.....	46
Figure (5.2): Code to Split Text into Paragraphs.....	46
Figure (5.3): Split the Text into Paragraphs.....	47
Figure (5.4): Code to Split Text into Sentences.....	47
Figure (5.5): Split the Text into Sentences.....	48
Figure (5.6): Code to Split Text into Words.....	48
Figure (5.7): Split the Text into Words.....	48
Figure (5.8): Removing Diacritics Marks.....	49
Figure (5.9): Placing Diacritics Marks.....	49
Figure (5.10): Extract the Original Word.....	50
Figure (5.11): Ontology Based Extraction.....	52
Figure (5.12): Code to Collect Global Terminology.....	53
Figure (5.13): Algorithm to Extract Definitions.....	54

Figure (5.14): Extracting the Enumerations.....	55
Figure (5.15): Extracting Verses of the Holy Quran.....	56
Figure (5.16): Extracting Comparisons.....	58
Figure (5.17): Extracting Cause-Effect.....	59
Figure (5.18): Implementation Flowchart.....	59
Figure (6.1): Results of Science Content	66
Figure (6.2): Results of History Content.....	67
Figure (6.3): Results of General Content.....	68
Figure (6.4): Curve of Science Text Results.....	70
Figure (6.5): Curve of History Text Results	71
Figure (6.6): Curve of General Text Results	72

List of Abbreviations

LO	Learning Objectives
KU	Knowledge Unit
KL	Knowledge Level
KLS	Knowledge Level Section
LOT	Learning objectives Template
NER	Named Entity Recognition
DTM	Degree of Template Matching
OT	Ontology Tree
IDE	Integrated Development Environment
ML	Machine Learning
MVC	Model-View-Controller
GT	Global Terminologies
LU	Learning Unit