جامعة الشرق الأوسط
MIDDLE EAST UNIVERSITY
Amman - Jordan

Automated Arabic Essay Grading System based on Support Vector Machine and Text Similarity Algorithm

نظام التصحيح الآلي للأسئلة المقالية في اللغة العربية باستخدام نموذج دعم المتجهات وخوارزمية تشابه النصوص

By:

Saeda Esmaile Odeh al-awaida

Supervisor:

Dr. Bassam Alshargabi

A Thesis Submitted in the Partial Fulfillment for the Requirements of the Master Degree in Computer Science

Department of Computer Science

Faculty of Information Technology

Middle East University

June, 2019

Authorization

I, **Saeda Esmaile Alawaida**, hereby authorize Middle East University to supply copies of my thesis to Libraries, organizations or individuals when required.

Name: Saeda Esmaile Alawaida.

Date: 03 / 06 / 2019

Signature: saeda

Thesis Committee Decision

This thesis titled "**Automated Arabic Essay Grading System based on Support Vector Machine and Text Similarity Algorithm**". Was successfully defended and approved on: 03 / 06 / 2019.

Thesis Committee Decision Members:

Name	Position	Workplace	Signature
Dr. Bassam Al-Shargabi	Supervisor	Middle East University	…………………
Dr. Mudhafar Al-jarrah	Head of the Committee and Internal Examiner	Middle East University	…………………
Dr. Mohammed Otair	External examiner	Amman Arab University	…………………

Acknowledgment

(اللهم لك الحمد والشكر كما ينبغي لجلال وجهك الكريم وعظيم سلطانك وعلو مكانك)

First, I would like to thank Allah for the strength and patience he had given me to finish this work. This work could not have been achieved without having faith that Allah is there to support and help me. May he bless everyone who was there for me during my studying period.

My words cannot describe how grateful I am to Dr. Bassam Alshargabi, whose recommendations, devotion, advocacy, patience, encouragement, and support have led me to achieve this work. I cannot express how lucky I am having him to supervise my thesis. In addition, I would like to express my deepest gratitude to all the respectable lecturers at the Faculty of Information Technology, Middle East University.

The Researcher

Saeda al-awaida

بسم الله الرحمن الرحيم

"وقل ربي زدني علما"

Dedication

This thesis is dedicated to my whole family;

Especial thanks to my precious Father and my one and only my Mother, who always proud of me and supported me in every step of my life, no words can describe what they have done for me, thank you for your endless love.

My brothers Mohammed , Ahmad ,Ibrahim, Mahmoud and **my sisters Samah, Amane, Amal and Eman who** are one part of my life.

My headmaster Mariam Rawahna and best friends who always been there for me during difficult and stressful times.

My friends, who supported me with their nice words, who were the cause of my happiness during my last days at university, and who loved them and will stay in my heart.

Table of Contents

List of Figures

List of tables

Table of Abbreviations

Abbreviations	**Meaning**
AEG	Automated Essay Grading
AES	Automated Essay Scoring
AAES	Arabic Automated Essay Grading
SVM	Support Vector Machine
TSA	Text Similarity Algorithms
OAO	One Against One
OAA	One Against All
NB	Naïve Bias
KNN	K Nearest Neighbors
EG	Essay Grading
PEG	Project Essay Grading
IEG	Intelligent Essay Grading
ETS	Educational Testing Service
E-rater	Electronic Essay Rater
NLP	Natural Language Processing
F-score	Fisher Score
LSA	Latent Semantic Analysis
RST	Rhetorical Statement Theory
VS	Vector Space
API	Application Project Interface
ASAP	Automated Student Assessment Prize

Automated Arabic Essays Grading System based on Support Vector Machine and Text Similarity Algorithm

By: Saeda Esmaile odeh Al-awaida

Supervisor: Dr. Bassam Al-shargabi

Abstract

The Automated Essay Grading (AEG) used in universities, companies, and schools which using computer technology capability to improve the grading system to overcome cost, time and teacher effort in correcting the student's paper. The Arabic Essay Grading system is wide spread over the world because they play critically in education technologies. AEG system applied for multiple languages such as (English, French, Bahasa, Hebrew, Malay, Chinese, Japanese, and Swedish). Therefore , this is thesis, focus on Arabic Essay Grading on the Arabic language, as there are many the techniques used in automated Arabic essay grading such as natural language processing and machine learning. Due to the lack of research on Arabic language AEG, this thesis introduced Arabic automated essay grading system consists of two main processes: firstly, Applying on Arabic WordNet to all possible or related word in meaning then select features based on support vector machine after the preprocessing step. Secondly process, evaluate electronic student essays according to previously determined answer models to find out the similarity degree using cosine similarity algorithm. According to the experimental result, reveal that the proposed system improves the performance of Arabic essay-grading as compared to human scoring.

Keywords: Automated Essay Grading, Support Vector Machine, Arabic WordNet, Cosine Similarity Algorithm.

نظام التصحيح الآلي للأسئلة المقالية في اللغة العربية باستخدام نموذج دعم المتجهات وخوارزمية تشابه النصوص

إعداد: سائدة اسماعيل العوايدة

إشراف: الدكتور بسام الشرجبي

الملخص

يتم استخدام نظام التصحيح الآلي (AEG) في الجامعات والشركات والمدارس التي تستخدم امكانيات تكنولوجيا المعلومات لتحسين نظام التصحيح الآلي للتغلب على التكلفة والوقت وجهد المعلم في تصحيح ورق اختبارات الطلاب. (AEG)انتشر على نطاق واسع في جميع أنحاء العالم لأنه ذو أهمية في تقنيات التعليم حيث الذي تم تطبيقه لعدة لغات مثل (الإنجليزية، الفرنسية، العبرية، الملايو، الصينية، اليابانية، والسويدية). ونظرا لقلة الأبحاث الخاصة بالتصحيح الالي للاسئلة المقالية في اللغة العربية تتناول هذه الرسالة هذا الموضوع وتطبيقاته في الأسئلة المقالية باللغة العربية، حيث تم استخدام العديد من التقنيات في تصحيح الاسئلة المقالية تلقائيا باللغة العربية مثل معالجة اللغة الطبيعية والتعلم الآلي وغيرها. وفي هذه الرسالة نقدم نظام تصحيح آلي للأسئلة المقالية خاص بالنصوص المكتوبة باللغة العربية، حيث يتكون من جزيئين رئيسيتين. أولاً، تم تطبيق WordNet باللغة العربية من اجل تحديد المرادفات والكلمات المشابهة لكل من الكلمات التي ذكرت في اجابة الطالب او الاجابات النموذجية، ثم نحدد الميزات باستخدام (SVM) بعد عملية المعالجة المسبقة. ثانياً، يتم تقييم اجابات الطلاب وفقًا لنماذج الإجابات المحددة السابقة لمعرفة درجة التشابه باستخدام خوارزمية التشابه المستخدمة (جيب التمام) واثبتت النتائج التجريبية في النظام المقترح عن تحسن في أداء دقه النظام المقترح.

الكلمات المفتاحية: تصحيح المقال الآلي، متجه دعم الالة (SVM)، WordNet باللغة العربية، خوارزمية التشابه (جيب التمام).

Chapter One

Background and the Study Importance

1.1 Research Context

This thesis focuses on automatic grading for the Arabic essay questions with a score comparable to human score by using a support vector machine to features selection and text similarity algorithms to find the score.

1.2 Background

Automated Essay Grading (AEG) continues attracting the attention of public schools, universities, testing companies, researchers and educators. A number of studies have been conducted to assess the accuracy and reliability of the AEG systems (Dikli, 2006). Furthermore, there were several AEG studies reported high matching rates between AEG systems and human raters with different techniques such as latent semantic analysis(LSA), support vector machine (SVM) and text similarity algorithms and combine between these technique (Zhang, 2010.). The vision of having effective algorithms to score student essays should be appealing to the teacher, test, and research scientist.65tg Teachers freed of the burden of reading and hand-scoring maybe hundreds of student papers and consequently, would be more likely to assign written questions and probe for the deeper understanding student.

Test publishers would be able to score essays for less cost and conceivably provide higher quality assigned grades with using computer's special capabilities and techniques to improve AEG system to achieve more accurate results compared to traditional scoring using standard measures mean absolute error and Pearson correlation result.

Many of the techniques used in AEG, such techniques within the field of natural language processing and machine learning and latent semantic analysis used to grade student essays(Alsaleem, 2011)(Al-Jouie & Azmi, 2017)(Suresh & Jha, 2018).

In this thesis, SVM technique and text similarity algorithms is used to extract feature and measure the percent of similar between model answer and student answer to find the proper score.

1.3 Definitions

1.3.1 Automated Essay Grading (AEG)

Automated Essay Grading (AEG) is a technique used to grade student essay without the direct participation of human which automatically evaluate the score or grade of a written essay to overcome time, cost and reliability. AEG systems motivated to develop solutions for assisting teachers in grading essays in an efficient and effective manner (Surya, et al., 2018).

Most AEG systems have implicitly or explicitly treated as a text classification problem, utilizing a number of techniques within the field of natural language processing and machine learning.

AEG system mechanism contained two stages: firstly, preprocess of the texts that are making the texts useful for further analysis and process after a collection of student texts in their text corpus forms inputted into the AEG system.

Preprocess techniques include stripping the texts of white space and removing certain characters such as punctuation, and remove any character from another language, and splitting text sequences into pieces, referred to as tokens.

Other methods employed in the preprocessing stage such as : tokenize, normalization ,stop word removal and stemming will be illustrated in the next sections in more details.

The second stage typically involves feature extraction, which is the process that maps the text sequences to a vector of measurable quantities,. It considered the most difficult part of the construction of an AEG system and it is a challenge for humans to take into account all the factors affecting in the grade. Furthermore, the effectiveness of the AEG system is constrained by the chosen features. (lilja, 2018)

1.3.2 Support Vector Machine (SVM)

Many of machine learning techniques such as Naïve Bayesian (NB), K- Nearest Neighbors (KNN), decision trees and Support Vector Machine (SVM) have been used in the scope of building an automatic essay grading. As for the SVM, the techniques investigated study in this thesis.

The SVM is a supervised classification algorithm, which was first proposed by Vladimir N. Vapid in 1963. It is based on the minimize errors in the classification (suresh & jha, 2018).

In this algorithm, plot each data item as a point in n-dimensional space with the value of each feature being the value of a particular coordinate. Then, perform classification by finding the hyperplane that differentiates the two classes.

The SVM has two classifiers model distinguish the first one (OAO) SVM classifier that classifies the data into two different classes (one class versus one class), another (OAA) SVM classifier that mean one class against multiple class which classifies the data into more than two classes. The multiple classes that mapping to multiple binary classifications. For example, if we have to classify data into M classes, which m is the number of classes, the number of classifiers define in equation 1.

$$N = \ {M(M-1)}/{2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

Where ***N:*** Number of classifier , ***M***: Number of class . For example, let us consider a typical two-class problem (class A, class B) Figure 1 shows a number of linear classifiers are possible for (class A, class B). SVM classification aims to find the linear classifier that maximizes the distance between (class A, class B) and the nearest data point of each class shown in Figure 2 (Sundaram Arun, 2015).

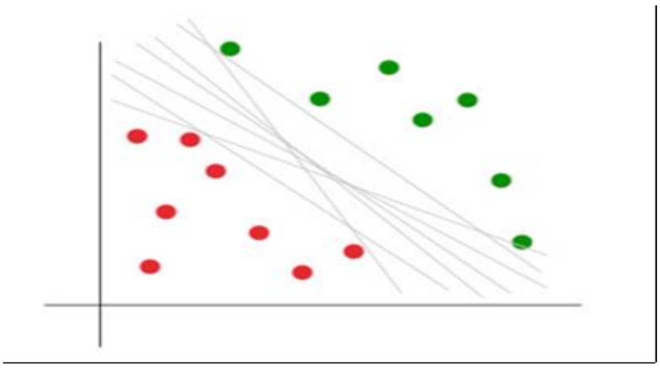

Figure 1: Number of possible classifiers (Sundaram Arun, 2015).

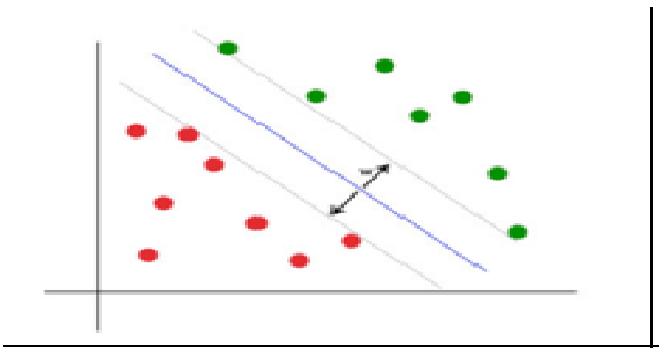

Figure 2: The nearest data point line of class (Sundaram Arun, 2015).

1.3.3 Text Similarity Algorithms (TSA)

Measuring the similarity between words, sentences, paragraph and documents are important for use in automated essay grading.

Text similarity contains three approaches: String-based similarities, Corpus-based similarities, Knowledge-based similarities, or a sample of combinations between them.

String-based similarities: partitioning them into two types character-based and term-based where these approaches measure the similarity by counting the number of different characters in these two sequences.

Corpus-Based similarities are a similarity measure that determines the similarity between words according to information gained from large texts that are used for language research.

The knowledge-based similarity is one of the semantic similarity measures that bases on identifying the degree of similarity between words using information derived from semantic network, Some of these were combined together to find the best performance was achieved by using a method that combines several similarity metrics into one.

1.4 Problem Statement

Recently, advances in electronic exam technologies have attracted considerable attention from universities and e-learning-based educational schools, helping e-learning to meet the needs of teachers and learners.

The traditional correction process, needs educational cadres and relatively high cost of money and a great time in sorting and checking student results, so, an automatic grading systems would help the teacher in cost and time.

Automated Arabic essay grading is still at the beginning, most of the methodologies used still do not achieve the accuracy required to achieve high precision in the correction process.

1.5 Question of the study

What the improvement in the automated grading system in terms of accuracy when using support vector machine and text similarity algorithms?

1.6 Objectives of the study

The main objectives of this study are:

1- Proposing an Arabic automated grading model that is based on support vector machine and text similarity algorithms.
2- Validating the proposed model in terms of accuracy.

1.7 Motivation

The main significance of the proposed technique is to enhance the accuracy of an automated grading system that based on composed of support vector machine and text similarity algorithms, as we need a system to generate scores for the student essay questions with a better accuracy that matches the human scoring.

1.8 Contribution

To help the teacher, to automatically score student essays freed of the burden of reading and hand scoring where hundreds of student papers and consequently would be more likely to assign written questions and probe for the deeper understanding student.

The system will be able to score essays for less cost and conceivably provide higher quality assigned grades with using computer's special capabilities to improve AEG system to achieve more accurate results compared to traditional scoring.

1.9 Scope and limitations

The scope of this study is to design an automated essay grading model based on SVM and text similarities approach. This study will be limited only for the grading essays written in the Arabic language.

1.10 Thesis outlines

Chapter 2: discuss the literature review and related work.

Chapter 3: presents the proposed technique.

Chapter 4: discuss the experiments work and result.

Chapter 5: will discuss the conclusion and future work.

Chapter Two

Literature Review and Related Work

2.1 Introduction

This chapter presents an overview of the concepts and main topics of Arabic automated essay grading, also presents the definition of Arabic automated essay grading including its history, and definition of support vector machine. This chapter will also define the text similarity algorithm and some types of it.

2.2 Automated Grading System

Based on the type examination as it is divided into two types : multiple choices and essay systems. Multiple choices systems are easy to implement but difficult to measure student-understanding courses that require asking students essay questions. Nevertheless, essays demand a better-measured depth level of understanding for the student (Rababah & Al-Taani, 2017). Accordingly, an essay grading (EG) through using specific computation technologies used to score essay questions by a number of studies conducted to assess the accuracy and reliability of the AEG system for multiple languages such as (English, Arabic...). From another side, the AEG have different systems will distinguish in the next section.

2.3 Automated Essay Grading Systems

Four types of AES system is used in testing universities and schools, The first automated essay scorer was developed by Ellis Page in 1966 with his Project Essay Grader (PEG) which concern to measure the quality of essay refer to writing construct such as word length, essay length, punctuation and soon (Hutchison, 2011). PEG system results were predicted score is nearest to human score this best advantage of the system. Later PEG was modified in several aspects in the 1990s; it incorporated special collections and classification schemes. Another system that Intelligent Essay Assessor (IEA) which they scored an essay using semantic text analysis method called latent semantic analysis which text is presented as matrix which rows in the matrix stand for a words and columns stands for context, and each cell included the word frequency, each cell frequency consider feature denoted a degree to which the word type carries information in the domain.

Electronic Essay Rater (E-rater) system was developed by the Educational Testing Service (ETS) to evaluate the score for an essay by identifying linguistic features (lexical and syntactic).

IntelMetric system is using a blend of artificially intelligent, natural language processing and statistical technology) which concern to measure the quality of essay refer to write construct and content of the text(Hutchison, 2011).

Table 1: provide a summary for the AES systems and their main approaches.

Table 1: Arabic Essay Scoring System

AES System	Developer	Technique	Main Focus
PEG	Page (1966)	Statistical	**Style**
IEA	Land Auer, Foltz, & Latham (1997)	LSA	**Content**
E-rater	ETS development team (Burstein, et al.,1998)	NLP	**Style and content**
IntelliMetric	Vantage Learning (Elliot, et al., 1998)	NLP	**Style and content**
BETSY	Rudner (2002)	Bayesian text classification	**Style and content**

2.4 Text Similarities

Text similarities are defined as the distance between words, sentences, paragraph, and documents based on the likeness of their meaning or lexically and semantically. Three approaches of text similarity String-based, Corpus-based, and Knowledge-based similarities. String-based operates on string sequence and character composition while character based on the distance between characters. Corpus similarity measures similar words according to information gained from large corpora. Knowledge-based similarity also measures the similarities between words and information derived from WordNet. In this thesis, the cosine similarity to grade the student answer is used.

2.5 Arabic WordNet

Arabic WordNet is a useful knowledge-based tool for several semantic similarity measures created in 2006 then had extended in 2015. It used in many natural language processing applications.

Arabic WordNet is a lexical database for the Arabic language, which concerns the meaning of words, rather than forms, words are semantically similar.

Also it lexical resources containing not only words of the targeted language but also synsets and semantic relations between them such as synonymy, meronymy, and antonym which Synsets are groups of words that each can substitute others in a sentence without changing its general meaning (Abouenour, Bouzoubaa, & Rosso, 2013).

This thesis uses Arabic WordNet to find all related words from student answer to give the answer of student a score. Students do not oppress in the mark because he did not write the same model answer exactly.

2.6 Support Vector Machine

SVM is a machine learning technique using for classification and features extraction which is features extraction methods of creating combinations of the variables to get around the problem while still describing the data with effective accuracy and it used to reduce the number of features from all input features which selecting useful features to perform the best classification.

Many methods used SVM for feature extraction such as Fishr score, Gradient Algorithm, K-means, RelieF, and SVM-RFE. This thesis the Fisher score is used which learns in more detail in the next subsection.

2.7 Fisher Score for Feature Selection

Support vector machine is represented by sparse vector s under the vector space, where each word in the vocabulary is mapped to one coordinate axis. Used on data to train a linear classifier which is characterized by the normal to the hyperplane dividing positive and negative instances.

The aim of apply feature selection that to Pre-defining the number of highest scoring features to be included in a classifier by using the F-score technique.

F-score is a simple feature selection technique in SVM, Which measures the distinction between two classes (positive and negative), the value of F-score for each feature is computed in the following equation (gunes, polat, & yossunkaya, 2010):

$$F(i) = \frac{\left(\overline{x_i^{+}} - \overline{x_i}\right)^2 + \left(\overline{x_i^{-}} - \overline{x_i}\right)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}\left(x_{k,i}^{(+)} - \overline{x_i^{(+)}}\right)^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}\left(x_{k,i}^{(-)} - \overline{x_i^{(-)}}\right)^2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

Where ***K*** is a positive or negative instance, ***n*** is a number of feature, (***xi*** ¯**,** ***xi*** +) the average of I feature positive and negative dataset, ***ki*** : the feature of the ***i***th positive /negative instances.

After determining, the score for each feature then obtained threshold value by calculating the average of F-score for all features.

If the value of F-score is larger than the mean value of all f-score, so that feature is added to feature space otherwise if F-score value is less than the mean value of all F-score, the feature is removed from feature space.

The Fisher score used in this study to decided or selected the feature that affected in the score of student answer which determined the positive and negative according to related for an answer or not related, which the related to answer (positive) take it but the others(negative) ignore it.

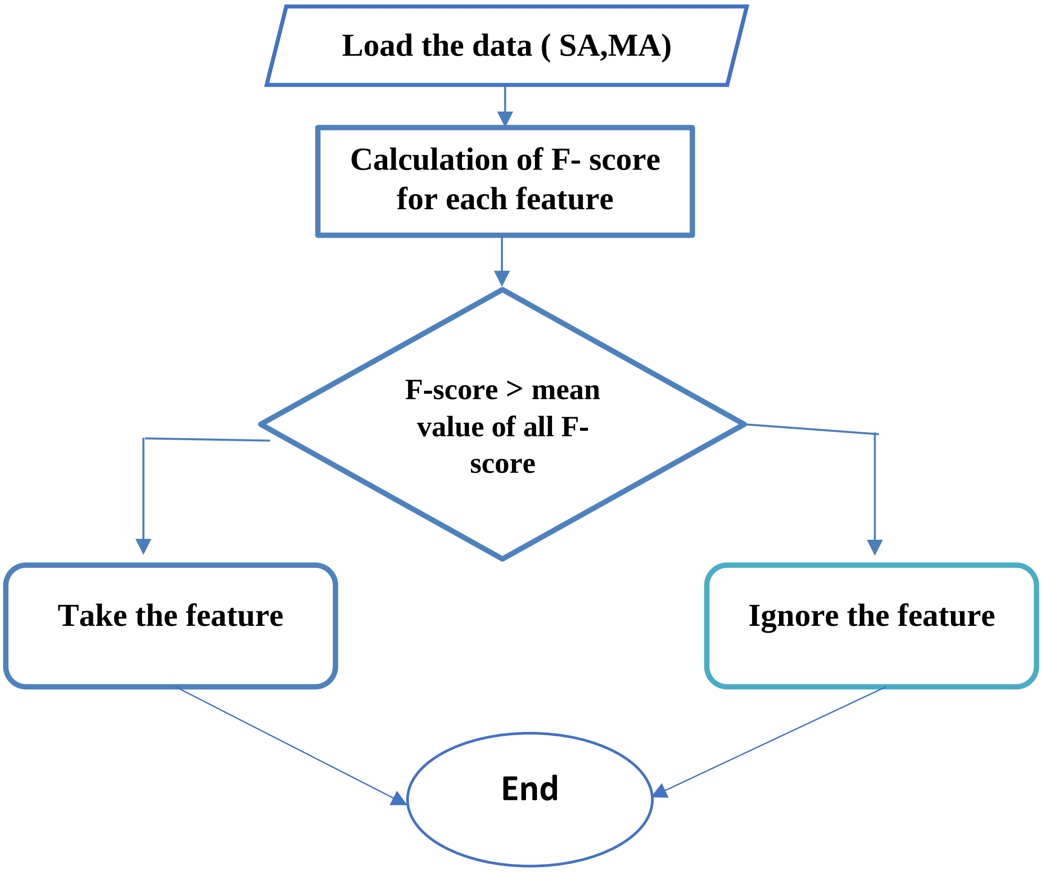

Figure 3: Flow chart of Fisher score feature selection

2.8 Related work

This section, describes the latest related work that presents different approaches for building AEG using SVM, LSA and text similarities follow as:

2.8.1 Latent Semantic Analysis (LSA)

A technique that uses statistics and natural language processing in information retrieval to get the semantic meaning in texts (content analysis of essay)

In (Refaat, Ewees, Eisa and Sallam, 2012) presented an automated assessor of Arabic free text answer based on LSA after unifying the form of letters, deleting the formatting, replacing synonyms, stemming and decreasing the number of stop words to be deleted (Refaat et. Al. ,2012 produced a matrix that better than the traditional form of LSA matrix, then using cosine similarity method to compare between the current answer and the model answers, then the large similarity ratio is taken to set a degree to current essay based on model answers degrees.

In (Alghamdi, Alkanhal , 2014) presented ' Abbir ' system for the Arabic language that was used LSA with some features such as word stemming, spelling mistake, the proportion of spelling mistake and word frequency to show that after a different experiment for automated essay scoring system the performance of very close to the human raters.

In (Mezher and N. Omar ,2016) proposed a modified LSA for automatic essay scoring using Arabic essay answers, a hybrid method of syntactic feature and LSA is based on Bag-of-words .after preprocessing create a matrix then apply cosine to define similarity, Results noted that syntactic feature improves the accuracy.This thesis use Arabic WordNet to apply the meaning features.

In (Al-jouie, Azmi, 2017) presented a hybrid method LSA and rhetorical structure theory for automated Arabic essay scoring this hybrid applies LSA for the semantic analysis of the essay, and the RST to assess the cohesion and the writing style of the essay. They assign 50% of the total score on the cohesion of the essay, 40% for writing style and the remaining 10% for spelling mistakes. After tested the system on the different school to achieve sufficient accuracy.

2.8.2 Support vector machine

This section, presents the approaches for building AEG that rely on the use of SVM for the feature selection process:

In (Gharib, Habib, Fayed, 2009) they applied multi-classifier such SVM, K-NN and Bayes in classifying documentation text in Arabic language and compare between them, using dataset from Aljazeera news web site and Al-Hayat website .according measures (recall, precision and F1) presented the result that SVM classifier significantly better outperform other classifiers in high dimensional feature spaces. accordingly, the SVM was used in our proposed model.

In (Alsaleem Saleh, 2011) presented a comparison between Naïve Bayesian method (NB) and SVM algorithm on different Arabic data sets. Using SVMs to define the hyperplane separating the space into two half-spaces with the maximum-margin to use in text classification. The results of all measures (F1, Recall, and Precision) against different Arabic text categorization data sets reveal that the SVM algorithm is better than the NB method in text classification. accordingly, the SVM was used in our proposed model.

In (Martinez, Dong Hong, Lee, 2013) proposed automated essay scoring system, firstly extract numerical features vector extracted from the text data of essays using support vector machine classifier, then construct a predictive model with extracted features and solve the multi-classification problem into multiple binary classifications to find the score between pairs of class. The results show that the performance of the proposed scoring system achieves accuracy near to teacher score.

In (R. Abbas, S.Al-qaza, 2014) suggested an Automated Arabic Essays Scoring (AAES) system in a web-based learning context based on the Vector Space Model (VSM).Two main processes approach ,firstly process extract the important information from essays, then apply support vector machine to find out the similarity degree between the previously written essays by the teacher and the essay written by the student after convert each essay to vector space, which using VS to matching terms in document after that we apply cosine similarity to find score of student answer. This thesis using SVM to extract feature from answers.

2.8.3 Text similarity algorithms

This section, presents the approaches for building AEG that rely on the use of text similarity approach for grading process:

In (Gomaa, Fahmy ,2013) suggested a short answer system written in the Arabic language to evaluate the student answer after they translated into English, to overcome the challenges in Arabic text, but some problem occurs through translation, such as a word in Arabic, not the same context structure and semantic translated. After that apply multiple similarity measures and combine between them to define the score of answer tested student. In this thesis, we directly apply similarity measure after extract feature without translated.

In (al-Jameel, James shea, Keeley, 2016) presented a survey for similarity approaches and challenges faced by the Arabic language. Three types of similarity were surveyed,

1- Lexical similarity based on (character similarity, statement similarity).
2- Semantic similarity based on (corpus similarity, knowledge-based similarity).
3- A hybrid similarity which combines between lexical and semantic.

The survey concluded that the cosine similarity measurement was used in many Arabic systems and compared to other lexical measurements the results show a more efficient performance. However, due to the different features in the Arabic language such as morphology and the semantic similarity using the lexical similarity approach is not reliable. It is not reliable because of the weakness of Arabic WordNet and Arabic Corpora. Because it combines more than one type of measurements which leads to the similarity being more robust, the hybrid similarity approach is considered as a promising approach with the Arabic language

In (Emad al Shalabi, 2016) presented a system for automated essay scoring of online exams in Arabic language that based on stemming technique in two approach heavy stemming and easy (light) stemming process and Levenshtein similarity measure to conduct question to check the efficiency of both mechanisms, where the light stemming is stopped removal of prefixes and suffixes, without attempting to identify the actual root of the word and heavy stemming referred to root-based stemming that removing prefixes and suffixes to extract the actual root of a word. after finding the stemming word the Levenshtein similarity measure done by giving each word a weight, then define the distance between every two words to find the score.

In (Shehab, Faroun, Rashad, 2018) presented a system based on the comparison of different text similarity algorithms for Arabic essay gradings such as string algorithm and corpus algorithm. They applied multiple similarity measures to find an effective solution for article grading systems. They used the N-gram approach has many advantages, such as simplicity; it is more reliable for noisy data such as misspellings and grammatical errors; and it outputs more Ngrams in given strings than N-grams resulting from a Word-based approach, which leads to collecting a sufficient number of N-grams that are significant for measuring the similarity.

Table 2 invoked the previous related work briefly:

Table 2: Related work.

Papers	Year	Description
Gharib, Habib, Fayed	2009	Applied support vector machine (SVM) in classifying Arabic text documentation, and using SVM classifier to convert data to vector space in high dimensional feature spaces significantly outperforms the other classifiers after that apply cosine similarity to score answer
Alsaleem Saleh	2011	Investigated Naïve Bayesian method (NB) and (SVM) on different Arabic data sets. and the results against different Arabic text categorization data sets reveal that the SVM algorithm outperforms the NB with regards to all measures

		(F1, Recall and Precision) according to that we use SVM to extract feature.
(M. Refaat, A. Ewees, M. Eisa, A. Sallam)	2012	Presented a assess Arabic free text answer based on LSA after unifying the form of letters, deleting the formatting, replacing synonyms, stemming and decreasing the number of Stop Words to be deleted to produce a matrix that better than the traditional form of LSA matrix and using Cosine Similarity method to compare between the current answer and the model answers, then the similarity measure is taken. In this thesis, we are using SVM.
Gomaa, Fahmy	2013	Suggested a short answer system written in Arabic language, evaluated the student answer after they translated into English to overcome challenges in Arabic text then apply multiple similarity measures.in this thesis apply similarity directly without translate.
Martinez, Dong Hong, Lee	2013	Proposed automated essay scoring system to construct the automated essay scoring system, we first extract numerical features using SVM classifier from the text data of essays then extracted features then define similarity. In this thesis, we apply similarity measure after extract features.
R. Abbas, S.Al-Gaza	2014	Suggested an Automated Arabic Essays Scoring system in web-based learning context based on the Vector Space Model that consists of two main processes. Firstly, the process deals with applying extract the important information from essays,

		then SVM is applied after convert information extraction as a vector space to find out the similarity using cosine measure.
Alghamdi, Alkanhal,	2014	Presented 'Abbir' system for the Arabic language that was used LSA with some features such as word stemming, spelling mistake, the proportion of spelling mistake and word frequency to show that after a different experiment for automated essay scorer the performance very close to the human raters.
Emad al Shalabi	2016	Presented a system for online exams in the Arabic language of automated essay scoring that based on stemming techniques (with heavy stemming and light stemming) and Levenshtein similarity. But in the proposed model use cosine similarity.
al-Jameel, James shea, Keeley	2016	Presented a survey paper for similarity approaches and challenging faced by the Arabic language, three types of similarity. Firstly that lexical similarity based on (character similarity, statement similarity), secondary semantic similarity based on (corpus similarity, knowledge-based similarity), finally that hybrid similarity which combines between lexical and semantic the result that cosine similarity more efficient in the Arabic language which used in term and knowledge similarity in the proposed model.
Mezher and N. Omar	2016	Proposed a modified LSA for automatic essay scoring using Arabic essay answers, a hybrid method of syntactic feature and LSA .they find a syntactic feature that concerned to pag

		of words, which the research concern to solve drawbacks in LSA. After preprocessing create a matrix then apply cosine to define similarity, Results noted that syntactic feature improves the accuracy, in this thesis we use Arabic wordnet to apply the meaning features.
jouie, Azmi	2017	Presented a hybrid method LSA and rhetorical structure theory for automated Arabic essay scoring this hybrid applies LSA for the semantic analysis of the essay, and the RST to assess the cohesion and the writing style of the essay. after testing the system on the different school to achieve sufficient accuracy.
Shehab, Faroun, Rashad	2018	presented a system based on comparison of different algorithms for Arabic automated assay system such as string algorithms and corpus algorithms, and they applied multiple similarity measures to find an effective solution for article grading systems, they achieved that N-gram better resulted than the other types of measures.

Chapter Three

Proposed methodology

3.1 Methodology Technique

The methodology approach used in this thesis is experimental to validate the result of proposed technique. The proposed technique proposes an essay-grading model to enhance accuracy scoring of student exams to match traditional scoring by using SVM to extract feature from text answer and similarity measure to define the score of student answer. The experimental work will use a dataset of questions and three levels of the answer will illustrate in more detail in next chapter. The results evaluated by accuracy measures (Pearson Correlation Result and Mean Absolute Error Value).

3.2 Outline of the proposed technique

The proposed model developed to enhance the accuracy of grading in essay exams. We use a dataset (corpus) that created for testing the model.

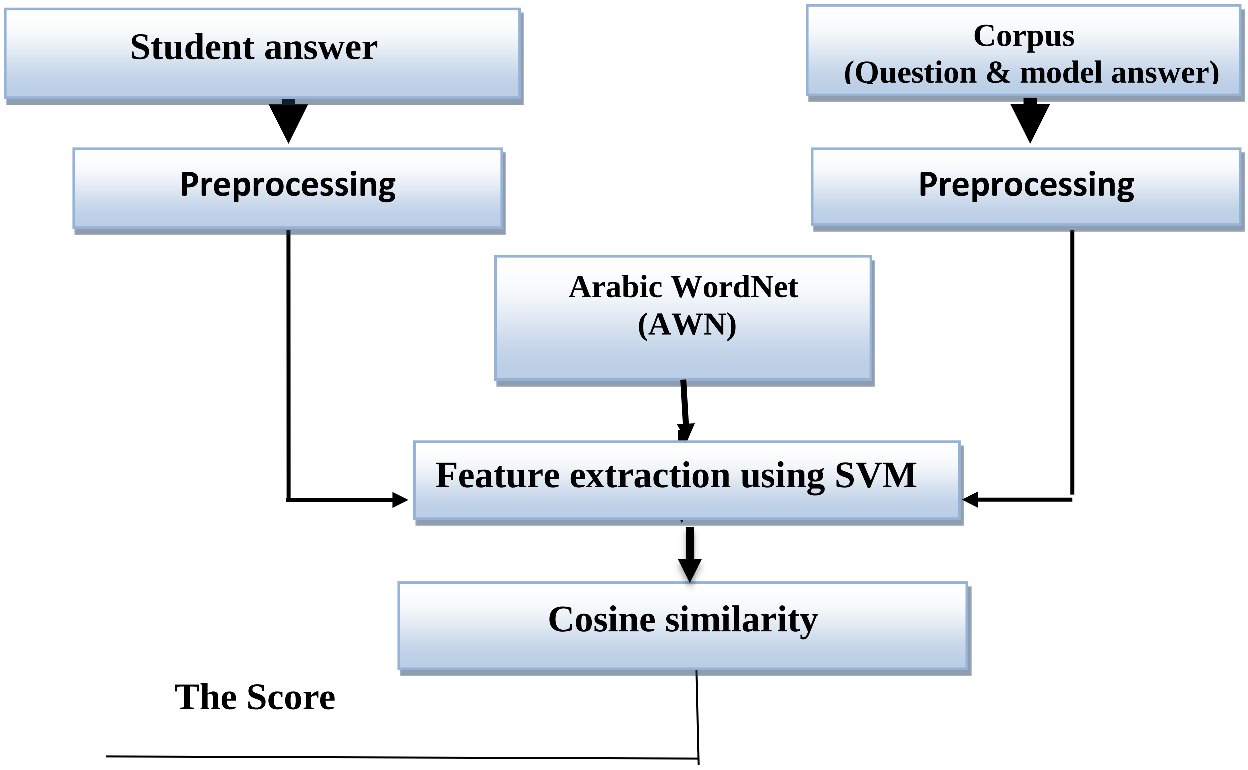

Figure 4: The Block Diagram of the Proposed technique.

3.3 Preprocessing

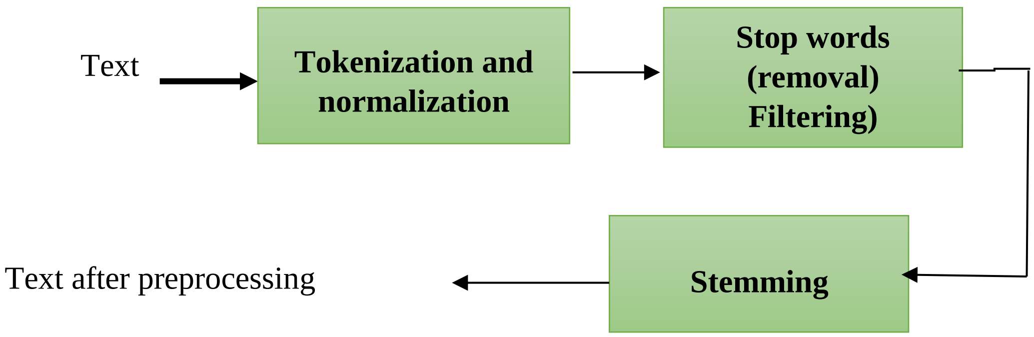

Figure 5: The Block Diagram of the preprocessing.

3.3.1 Tokenization and normalization process

Tokenization is a preprocessing step, which splits strings of student answer and model answer into smaller pieces. Text of student answer and model answer in the corpus is a sentence that separated by a stop mark such as ("،", ".", "؟" or "!"). Can be tokenized into sentences, then sentences can be tokenized into words, etc.

Text "مجموعة من الشبكات المتصلة مع بعضها البعض"

Token as "مجموعة" ,"من", ,"الشبكات" "البعض", "بعضها", مع " , "المتصلة"

The normalization process is the process of transforming character and words into a single form. Normalizing text before processing it allows for separation of concerns since input guaranteed to be consistent before operations performed on it. Text normalization requires being aware of what type of text is to normalize and how it is to be processed.

3.3.2 Arabic Stop words

The stop words can define as words that do not have any remarkable importance or any word that do not give any importance and meaning in finding text classification, so they removed these words from the text. After converting the input Arabic text to a list of tokens, then inputted to the next stage which they stop words removal will be listed in the dictionary to remove it from tokens output. For example , " كلما " , "حيث" , "الى" "مع" اللذان"، " من"" "في", "فوق", "تحت", "هذا", "هذه", "الذي, "التي", "نحن", "اولئك", "عادة" and so on. These words will remove from text to assure useful classification. In this research, they list the stop word in the dictionary to use in classification.

Token as " مجموعة","من","الشبكات","المتصلة", "مع “, "بعضها", "البعض"

After applying the stop word removal output as ,"المتصلة" ,"الشبكات" " مجموعة"

3.3.3 Stemming

A stem is a procedure which retrieves the word to basic root, by processing removing all words prefixes and suffixes and infixes Lemmatization is closely related to stemming which extracts the base root of words. It creates an actual dictionary for words.

Different types of stemmer used in Arabic text classification in this research we use ISRI Arabic Stemmer is used to extract the roots of the word.

ISRI is oriented towards finding the minimal representation of a word, which is used for information retrieval after normalizing the input word, removing diacritics, and non-related Arabic characters, stemming process, stopped when the remaining length of the input word is three or fewer characters (El-Defrawy , Belal , & Elsonbaty, 2015)(Al-Shargabi, Olayah, Al-romimah 2011) Al-Shargabi, Al-romimah, Olayah 2011). Example the word "الذهاب" input to stemming then the output is the third root is "ذهب". Previous stage stop word removal output as: ,"المتصلة" ,"الشبكات"," مجموعة" The stemming process output as: " وصل ","شبك" ,"جمع".

3.4 WordNet (Arabic WordNet)

WordNet is a lexical database, It groups words into sets of synonyms called synsets when use for Arabic language, also records a number of relations among these synonym sets, which it find a lexical resource offers broad coverage of the general lexicon to each word in student answer that extracted from the previous stage to define all the words that have near to meaning. Used in this thesis to find all the words that are synonymous with the student's answer to increasing the likelihood of the correct answer to the student which was used after the preprocessing step.

For example, the word "جمع"defines all the words, which possibly have it or near meaning according to a list of data that stored previously in the system. Also, the words "وصل" ,"شبك"make the same process to define the possible alternative words which the same and related.

3.5 Support Vector Machine

SVM is a machine-learning method that used for Text Classification, creating a feature space to use some scoring function to rank each feature, and then choose the best k features by using F-score method, which this method used to extract feature in SVM. In case, the data is nonlinearly separable, SVM makes the data linearly separable using kernel functions. A kernel function maps the input data patterns to some high dimensional space according to the text, to make the points linearly separable in high dimensional space.

3.5.1 Feature space

The features and samples from the student answer defined as each individual token occurrence treated as a feature. Then the vector of the entire token for a given document considered a multivariate sample. The SVM that maps feature to a higher dimensional space and tries to separate the classes.

Usually, a text vector spans your vocabulary size. Let's consider we have four texts in the corpus:

Text 1: **مجموعة من الحواسيب ترتبط فيما بينها بوساطة خطوط اتصال لها القدرة على نقل البيانات والمشاركة في المعدات**

Text 2: **مجموعة من الحواسيب ترتبط فيما بينها بوساطة خطوط اتصال**

Text 3: **مجموعة من الحواسيب ترتبط فيما بينها**

Text 4: [ ]

After some preliminary filtering, stop word removal and stemming, you obtain.

Text 1: جمع – حسب- ربط – وسط –وصل- قدر- نقل – بين – شرك –عدد

Text 2: جمع – حسب- ربط – وسط –وصل

Text 3: جمع – حسب- ربط

Text 4: [ ]

Let's have a look at the total vocabulary now: -

جمع – حسب- ربط – وسط –وصل- قدر- نقل – بين – شرك –عدد

you have got 10 words. sorted them in lexicographical order. Your vocabulary size is 10. Therefore, the vector will have 10 dimensions...

جمع – حسب- ربط – وسط –وصل- قدر- نقل – بين – شرك –عدد

Let's fit our documents in the transform:

Text 1: [1,1,1,1,1,1,1,1,1,1]

Text 2: [0,0,0,0,0,1,1,1,1,1]

Text 3: [0,0,0,0,0,0,0,1,1,1]

Text 4: [0,0,0,0,0,0,0,0,0,0]

This is bag of words model. fill in each space of the vector based on whether the corresponding word in the vocabulary exists or not.

The student answer inputted to preprocessing to filter and remove stop word and stem then represent a list of words in student answer after that created a dimensional of a word to decide which class is more significant to student answer to give a score. After selecting the most significant terms in the super vector, each answer text is represented as a weighted vector of the terms found in the vector space as shown in figure 6.

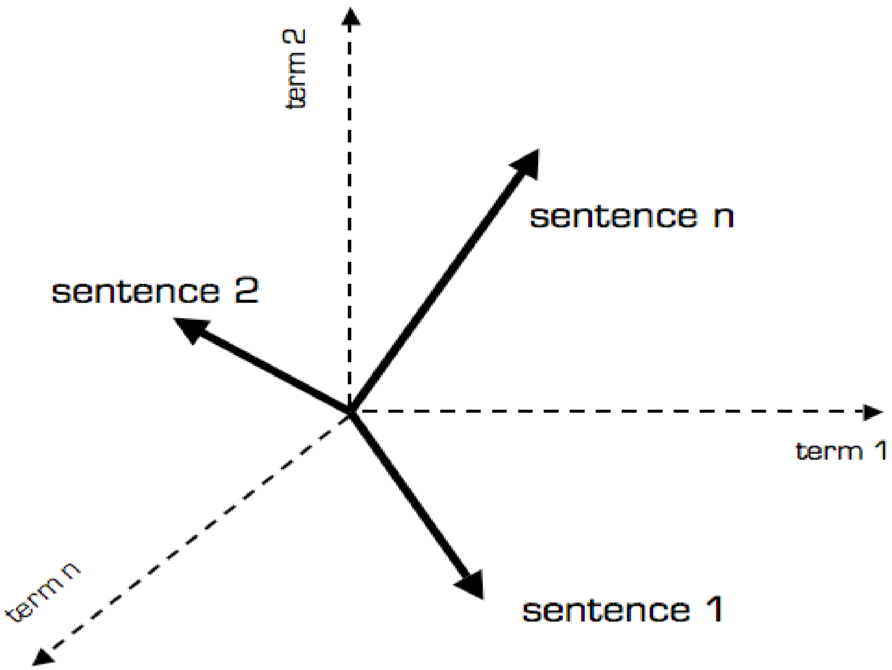

Figure 6: Example of implement vector space.

3.5.1.1 Term weight

After selecting the most significant term in the vector, each model answer presented as a weighted vector of the term found in the super vector. Every word in the model answer is given a weight. There are used the term frequency/inverse document frequency (TFIDF) weighting scheme in this thesis. Which define the weight for each term using equation 4 that selected from answers (Fouad Gharib, Badieh Habib, & Taha Fayed, 2013):

$$w_{t,d=TF_{t,d}} . \log \frac{N}{DF_T} \quad \ldots\ldots 3$$

Where ***t***: denote to term, ***d***: denote to document, ***N***: denote to number of documents, ***df***: denote to the frequency of document.

TF-IDF is one of the most popular term weighting, Used for measuring the information content of the terms in the documents. The number of times a word appears in the document called TF and the number of documents in the corpus that contain the word is called IDF.

3.6 Text Similarity Approaches

The similarity is the measure of how many matches data between two documents (A, B). In addition, Similarity in data context usually described as a distance with dimensions representing features of the document. The similarity between words is a fundamental part of text similarity, which then used as a primary stage for sentence, paragraph and document similarities.

Different type of similarity measure that used in two way lexical and semantic, a lexical similar which has the same character sequence. Semantically similar if they have the same word, used in the same way, used in the same context and one is a type of another.

The similarity measure used to define the score of student answer .After select ranking answer to reducing the computation of numeric scores on document pairs; a baseline score function for this operation the cosine similarity between student answer and ranked answer representing the query and the document in a vector space model.

3.6.1 Cosine similarity

It measures the cosine of the angle between two vectors space. It can see as a comparison between documents on a normalized space because we are not taking into consideration only the weight of each word count of each document, but the angle between the documents. This is the cosine similarity formula. Cosine Similarity will generate a metric that says how related are two documents by looking at the angle instead of magnitude (rahutomo & kitasuka, 2012).

$$\vec{a}\,.\,\vec{b} = \|\vec{a}\|\ \|\vec{b}\|\ \cos\theta \quad \text{...4}$$

$$\cos\theta = \frac{\vec{a}\,.\vec{b}}{\|\vec{a}\|\ \|\vec{b}\|} \quad \text{... 5}$$

Where *a, b* is a vector space

$$\cos\theta = \frac{\sum_i w_{q,i} . w_{i,j}}{\sqrt{\sum_j w_j^2} . \sqrt{\sum_j w_{i,j}^2}} \quad \text{... 6}$$

Where $w = tf$, *i* is a document, *j is a* term , *Wij* :is the weight of term *j* in document *i.*

If had a vector pointing to a point near from another vector, they have a small angle measurement near zero that means high similarity between document ***a*** and ***b*** shown in figure 7 part one. while the angle between two vectors is 90, so cosine of the angle is 0 that means the two documents ***a*** and ***b*** don’t have any similarity (nonrelated) between them, shown in figure 7 part two. The last one shown in figure 7 part three the angle is 180, so the cosine of an angle is -1 which mean the two documents are the opposite of each other.

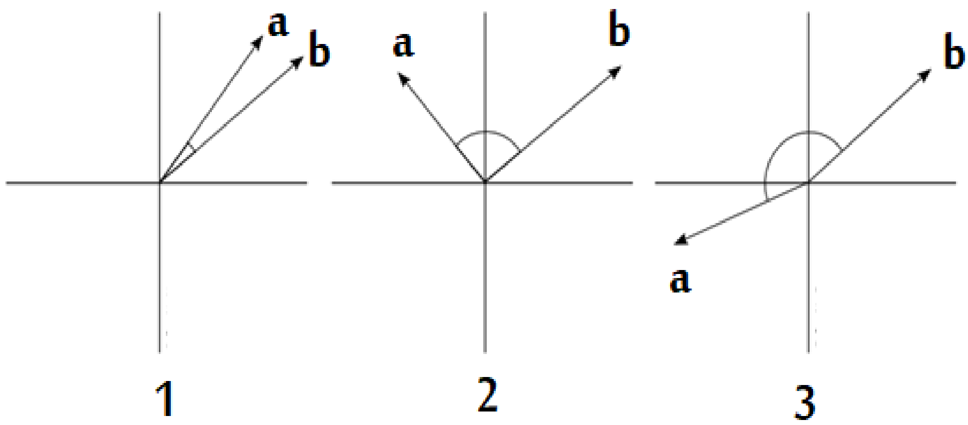

Figure 7: Cosine similarity values for document (***a* , *b***)

Chapter Four

Experimental Results and Discussion

4.1 Introduction

In this thesis, the proposed automated Arabic essay grading system used the SVM technique fisher score to extract features from text, then used cosine similarity to score student answer. The proposed technique was implemented in python (2.7) because it is more widely used in machine learning and support library for Arabic processing, also using models in python Django Application Program Interface (API) to create tables using class of model. Because it faster deployment and easy to perform. The experimental work included inputted student answer to system then preprocessing the text to take the keywords and vocabulary which most significant, to convert in vector space in n-diminution, the same process goes for model answers derived from dataset for the same question, after that we apply similarity measure to find score of the student answer .

4.2 Objective of the experimental work

The experimental work aimed to evaluate scoring accuracy using the proposed automated essay-grading model. The following objective is considered:

1- Creation of the dataset of question and model answer of the student.

2- Creation proposed model by python, preprocessing of student answer and stored model answers in the dataset, then apply the Arabic WordNet, then using support vector machine to extract features, finally using similarity measure to define the score of the answer.

3- Calculate the accuracy of the proposed model use Mean Absolute Error and Pearson Correlation Coefficient Results.

4.3 The collected dataset

The dataset created in MYSQL as CSV file, data collected from computer, science and social books from Allu'lu'a modern school, contain 40 questions and 3 classes of the answer. That built on the database style in Hewlett Foundation Automated student assessment prize. Used models in python Django (API) to create tables using class of model because this faster deployment and easy to perform. The first table for question data and table two for model answer for question. The parameters of the dataset are as follow:

Id: is a unique Id introduce for each answer.

Score: the score that was given by examiner to student answer of the particular data.

Answer-text: represent the answer given by students on the

Question Id: is a unique Id introduced for each question.

a sample of the dataset for questions shown as in figure 8:

id		correct_answer	Text	question_name
"السؤال الاول', 'عرف الانترنت', 'مجموعه من الاجهزة المتصلة مع بعضها البعض', '1'				
"\\السؤال الثاني ', 'ماذا يقصد بأمن المعلومات ', 'هي عمليه ابقاء المعلومات تحت سيطرتك الكامله', '4'				
"السؤال الثالث', 'ماذا نعني بمحركات البحث', 'هي برامج متخصصه في الشبكه الافتراضيه تستخدم للبحث عن المعلومات تساعد الباحث للحصول على المعلومات', '3'				
"السؤال الرابع', 'ما هي متطلبات الاتصال بالانترنت ', 'جهاز حاسوب , موديم , اشتراك من احد الشركات المزوده للخدمه و متصفح انترنت', '4'				
"السؤال الخامس', 'ماذا يقصد بمزود خدمه الانترنت', 'شركه تمكن المشترك من الحصول على الانترنت', '5'				
"السؤال السادس', 'ما هي وظيفه جهاز المودم', 'جهاز يربط بين الحاسوب وخط الهاتف', '6'				
"السؤال السابع', 'ما هي الخدمات الي تقدمها الحكومه الالكترونيه , وما فائدتها', 'تتكون الدارة الكهربائية المصباح', '7'				
"السؤال الثامن', "اذكر خدمات البريد الالكتروني', 'مشاركه الرسائل و الملفات والصور', '8'				
"السؤال التاسع', 'علل , احترام الحقوق المليكيه الفكريه عند استخدام الانترنت', "لان حقوق النشر ونسخ المواد الموجوده مملوكه لاشخاص اخرين و لا يحق لاحد ان يعيد نشرها", '9'				
"السؤال العاشر', 'عرف بروتوكول الانترنت', 'هي القواعد التي تتحكم في الاتصال و تبادل المعلومات للاجهزه المختلفه المرتبطه بالشبكه', '10'				
"السؤال الحادي عشر', 'عرف الصراع الاجتماعي ', 'عملية اجتماعية تؤدي الى هدم المجتمع', '11'				

Figure 8: Questions dataset

Then the sample dataset created by the researcher for model answers shown in the following figure 9, more details can be seen in appendix A.

```
id,score,Answer_Text,question_id
1,100,"مجموعه من الحواسيب ترتبط فيما بينها بواسطه خطوط اتصال لها القدره على مشاركه البيانات",١
2,75,"مجموعه من الحواسيب ترتبط مع بعضها بوساطه خطوط اتصال",١
3,25,"مجموعه من الحواسيب ترتبط مع بعضها البعض",١
4,100,"هي عمليه ابقاء المعلومات تحت سيطرتك الكامله ، اي عدم امكانيه الوصول اليها من اي شخص اخر",٢
5,75,"ابقاء المعلومات تحت سيطره شخص ",٢
6,25,"ابقاء المعلومات بحوزتك ومن دون تدخل",٢
7,100,"هي برامج متخصصه في الشبكه الافتراضيه تستخدم للبحث عن المعومات",٣
8,75,"تساعد الباحث للحصول على المعلومات",٣
9,25,"هي عباره عن برامج بحث في الانترنت",٣
10,100,"جهاز حاسوب ، موديم ، اشتراك من احد الشركات المزوده للخدمه ومتصفح انترنت",٤
11,75,"جهاز حاسوب شخصي ، او خط هاتف ،وبعض المتطلبات الاخرى",٤
12,25,"موديم وجهاز",٤
13,100,"هي الشركه التي توفر لعملائها امكانيه الوصول الى الانترنت عن طريق الاشتراك",٥
14,75,"شركه تمكن المشترك من الحصول على الانترنت ",٥
15,25,"امكانيه الاشتراك باللانترنت",٥
16,100,"جهاز يربط بين الحاسوب وخط الهاتف و يقوم بتحويل البيانات الصادره من جهاز الحاسوب الى اشارات موجيه تنتقل الى خط الهاتف",٦
17,75,"هو جهاز يستخدم لربط الحاسوب وخط الهاتف وقد يكون في الداخل او الخارج",٦
18,25,"جهاز يستخدم لربط الحاسوب وخط الهاتف",٦
19,100,"تقديم المعاملات الحكوميه و الخدمات العامه عبر الانترنت ، فائدتها تحسين الاداء الحكومي التقليدي وتقليل الوقت و التكلفه",٧
20,75,"تقديم الخدمات و المعاملات لتوفير الوقت",٧
21,25,"تقديم المعاملات و انهائها دون الحاجه الى اللجوء الى الذهاب الى الدائره الحكوميه",٧
22,100,"مشاركه الرسائل و الملفات والصور ",٨
23,75,"تبادل الرسائل بين مرسل و مستقبل",٨
24,25,"عمليه تبادل الرسائل بين اكثر من شخص",٨
25,100,"لان حقوق النشر ونسخ المواد الموجوده مملوكه لاشخاص اخرين ولا يحق لاحد ان يعيد نشرها",٩
26,75,"لانها ملك لاشخاص ولا يحق التعديل عليها",٩
27,25,"هذا يعتبر تدخل بخصوصيه الاشخاص",٩
```

Figure 9: Scoring dataset for model answer

4.4 Work Procedures

The experimental work consists of three main modules:

1- Preprocessing text

2- Feature selection

3- Similarity measure

4.4.1 Preprocessing text

In the total all of the data of student answer and model answer from the corpus, we are processing by python library that used for Natural Language Processing (NLP). The library is NLTK contains different text processing libraries for classification, tokenization, and stemming.

Firstly, after the text input the tokenization and normalization do to split the text to words then removing spaces and punctuation, also take the same shape for some character such as (أ, آ, إ) to standardize the process. Secondly, the stop word will be, remove that will illustration and explain in the next subsection.

For example, answer student:

مجموعة من الحواسيب والأدوات والمعدات ترتبط فيما بينها بوساطة خطوط اتصال

The result after tokenizing:

"مجموعة" " من " " الحواسيب " "الادوات" " ترتبط " " فيما " " " خطوط" " اتصال "

بينها" " "المعدات" " بوساطة "

4.4.1.1 Removing stop word

After splitting the text, the system found words do not have any meaning, which these words listed in the system this some it.

إذ، إليكن، إذا، إذما، إذن، أف، أقل، أنا، أكثر، التي، إلا ألا، الذي، اللاتي، اللائي، أم، أن، إما، إنه أنى، آها آه، بي، إليك، أولا أولئك، أوه، آيا، أي، أنتم، أنت، إن، اللتان، التيا، اللتين، اللذان، أنتن، اللذين، اللواتي، إنا، إلى، إليكما أما الذين أما، أو، أينم، ا إيه، بس، بعد، بكم، بعض، بك، بكم، بكما، بكن، أين، إي، أيها، إنما أنتما إليكم بماذا، بمن، بنا، به، بها بل، بلى، بما، أين، بهم، بهما، به، بين، تلك، تلكم، ثم

After apply stop word removal for previous the result that:

"بوساطة" "مجموعة" "الحواسيب " "المعدات" "الادوات" " ترتبط " "خطوط" "اتصال"

4.4.1.2 Stemming

Stem algorithm employed in a different task in information retrieval. The algorithm which used in stemming is IRIS stemming (Taghva et al., 2005) that applied on the normalized word then follows a series of decisions to remove possible prefixes and suffixes to give roots from words, stemming process should be stopped when the remaining length of the input word is three or fewer characters. In this work, we use the IRIS stemming which bring it from NLTK library.

After applying stem process on previous output the result that:

"جمع" "اداة" "عد" "حسب " " ربط " "وسط" "خط" "وصل"

4.4.2 Arabic WordNet

After extract root of each word in the student answer along with model answers, then apply the word net to define all possible or related word in meaning for each extracted from the text as shown in table 3:

Table 3: Example of WordNet

word	synonyms
وصل	بلغ, علم, نهى ....
وسط	قصد, جزع,قطع ....
ربط	وثق, شد ....
جمع	حقن, قرن, لصق, ألف, حفظ, حشد,.................

4.4.3 Feature extraction

A common method that is used to perform supervised learning to select features using Support Vector Machine (SVM) with the Fisher score approach, as shown in chapter two it used to select a feature from student answer and model answer in corpus to decided which positive or negative feature as shown table 4:

Student answer: "جمع" "اداة" "عد" "حسب" " ربط " "وسط" "خط" "وصل"

Table 4: Fisher score feature selection

Positive	"جمع" "حسب" "ربط" "وسط" "خط" "وصل"
Negative	"اداة" "عد"

In the last step, used cosine similarity to define the degree of similarity between the student answer and model answer after previous processing, to give a score of students answers.

4.5 Experiment work

In order to evaluate the proposed model effective in the automated essay grading system, this study carried out a comparative analysis of the impact of Arabic WordNet in automated essay grading. The experimental work divided into two-stage as follow:

4.5.1 Experiment one

This experiment implements the result of the proposed model without using Arabic WordNet, with a comparison between human score and automated score as shown in table 5:

Table 5: Result of the proposed technique without WordNet

Question(id)	human score	cosine result without WordNet
1	1	0.94
2	1	0.94
4	0.75	0.67
5	0	0
11	0.75	0.66
12	0.75	0.82
15	0.25	0.21
30	0.25	0.22
36	0.25	0.1
40	1	0.97

4.5.2Experiment two

The next section that implements the proposed model by using WordNet with the result shown in table 6:

Table 6: Result of the proposed technique using WordNet

Question(id)	human score	cosine result with WordNet
1	1	0.98
2	1	0.98
4	0.75	0.8
5	0	0
11	0.75	0.67
12	0.75	0.85
15	0.25	0.24
30	0.25	0.3
36	0.25	0.21
40	1	0.98

Table 7: Comparison result between cosine result with WordNet and without WordNet

Question(id)	cosine result with WordNet	cosine result without WordNet
1	0.98	0.94
2	0.98	0.94
4	0.8	0.67
5	0	0
11	0.67	0.66
12	0.85	0.82
15	0.24	0.21
30	0.3	0.22
36	0.21	0.1
40	0.98	0.97

The figure 10 gives a graphical description of the result of two part of the experiment.

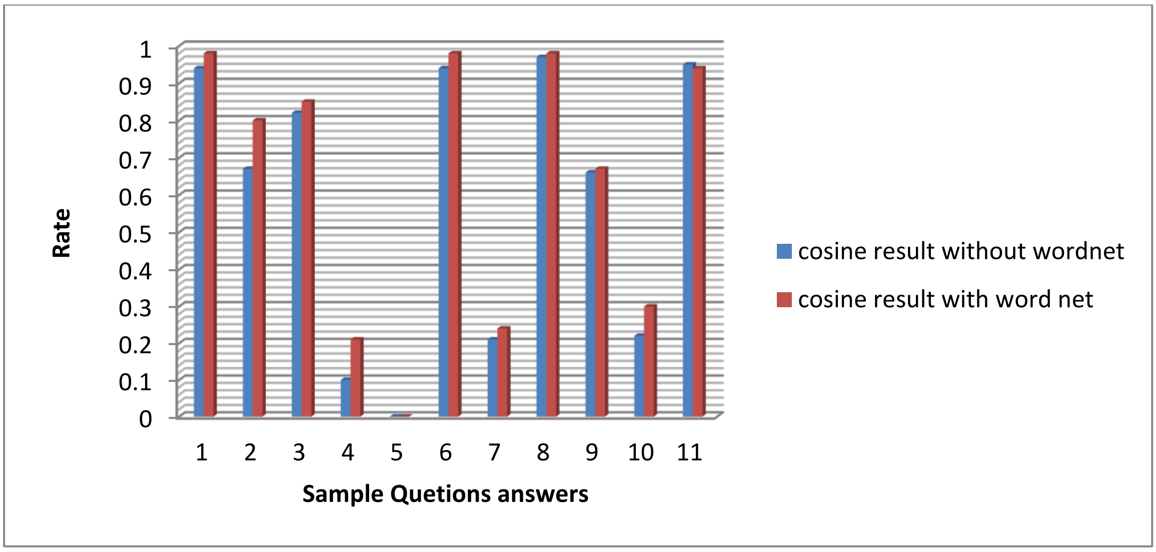

Figure 10: Rates of Cosine similarity with and without Arabic WordNet

The results of the automatic grading system when using Arabic WordNet are better than the results without the use of Arabic WordNet as shows in table 7 and figure 10 where it shows a difference in cosine result.

4.6 Evaluation Result

Evaluation of WordNet in Arabic automated essay grading performed by comparison between the human score and automated score for student answer using the mean absolute error value and the Pearson correlation coefficient:

4.6.1 Mean Absolute Error Value

The mean absolute error value is used to determine the accuracy of the proposed technique. Equation 7 used to derive the mean absolute error value:

$$MAE(\overline{X}) = \frac{\sum_{i=1}^{n}|x_i - y_i|}{n} \qquad \ldots\ldots 7$$

Where ***x***: human score, ***y***: automated score, ***n***: number of data test.

For mean absolute error for human score and automated essay score using cosine similarity with WordNet and without using WordNet, after measured listed in table 8

Table 8: Mean absolute error

	Cosine similarity with Arabic Word Net	Cosine similarity without Arabic Word Net
MEAN ABSOLUTE ERROR	0.117335572	0.120527628

The mean absolute error value between human score and AAEG using cosine similarity with AWN and without use WordNet were computed for our dataset as shown in table 7, an MAE of AAEG using cosine similarity with Arabic WordNet is 0.117 is less than MAE of AAEG using cosine similarity without using Arabic word net, so this result indicates that the proposed model will improve in the Arabic Automated Essay System with Arabic WordNet. As shown in figure 11.The improved accuracy of proposed technique for MAE with AWN as compared to MAE without AWN is computed in equation 8 :

$$\text{Enhancement } = \frac{MAE1 - MAE2}{MAE1} * 100\ \% \ldots\ldots 8$$

According to the values of MAE shown in table 8 and figure 11 , the enhanced accuracy is 2.648 % .

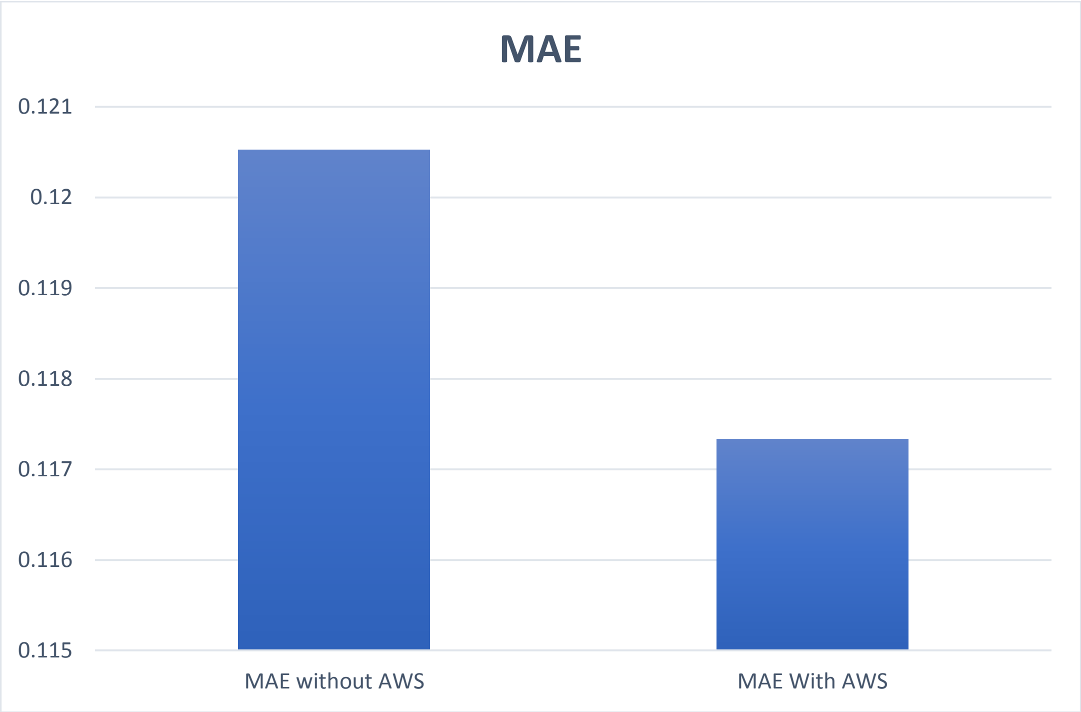

Figure 11: Mean Absolute Error Result

4.6.2 Pearson Correlation Result

Pearson Product Moment Correlation ***r*** that measure the strength between variables and relationship between them, equation 8 used to measure the correlation between the dependent variable and independent variable which in this thesis indicate to human score for the dependent variable and automated score for the independent variable.

$$r = \frac{\sum xy.\frac{\sum x.\sum y}{N}}{\sqrt{\sum x^2 - \frac{\sum x^2}{N}.\sum y^2 - \frac{\sum y^2}{N}.}} \ldots\ldots 9$$

Where ***x***: Human score***, y***: Automated score, ***N***: number of question test.

Table 9: Pearson Correlation Result

	Cosine similarity with AWN	Cosine similarity without AWN
Pearson Correlation Result	0.990227853	0.989475216

Equation 8 represents the Pearson correlation coefficient formula, the valid result for ***r*** lies between -1 and +1. If the result lies between 0 and 1, it shows there is a positive correlation that is ***X*** increases as ***Y*** increases. If ***r*** *= 1*, it shows that the result is perfect positive. If ***r*** is between 0.5 and 1, it shows a high positive correlation, when ***r*** is between 0 and 0.49, it exhibits a low positive correlation. When ***r*** = -1, it shows a perfect negative correlation that is the rate at which the dependent variable increases is exactly equal to the rate at which the independent variable decreases. When ***r*** is between -0.5 and 0, it shows a weak negative correlation, when r is between -0.49 and -1, it exhibits a strong negative correlation.

As shown in table 9 the Pearson correlation result for the proposed technique compared to human score ***r*** is between 0.5 and 1, it shows a high positive correlation that represented having the best correlation magnitude. In addition, the figure 12 give a graphical description of the result indicate that Arabic automated essay grading using cosine similarity with *AWN* in this study significantly correlate to human score.

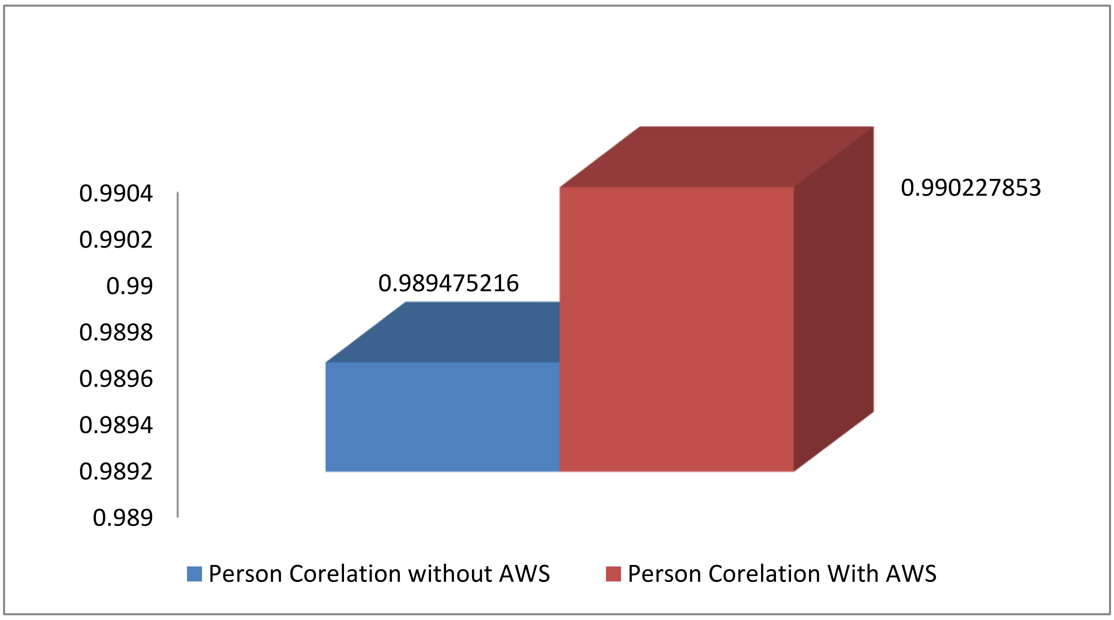

Figure 12: Pearson Correlation Result

Chapter Five

Conclusion and Future Work

5.1 Conclusion

This thesis presented the Automated Arabic essay grading model to achieve the accuracy with using support vector machine to select a feature from student answer and model answer, after using Arabic WordNet to get more choices for student answer. Finally, we use cosine similarity to define the score of student answer. The focus of this work was on enhancing the accuracy of Automated Essay System to match human score by adding Arabic WordNet. The dataset created which contain 40 questions with 120 answer model according to helmet ASAP form Kaggle datasets, the AAEG proposed in this thesis, as experimental results shows that the AAEG with using the WordNet is better in terms of accuracy compared AAEG without using WordNet according to mean absolute error value and Pearson correlation.

5.2 Future Work

In the research fields, there is not complete research, but each research work can provide new ideas for another work. Based on the outcome of the present research, the following ideas are suggested for future work:

1- Using machine learning and neural network models to enhance accuracy.

2- Using big dataset to implement the system.

3- Add word-embedding technique.

Reference

Al-Jouie, M. F., & Azmi, A. M. (2017). Automated Evaluation of School Children Essays in Arabic. *Procedia Computer Science*, *117*, 19–22.

Alghamdi, M., Alkanhal, M., Al-Badrashiny, M., Al-Qabbany, A., Areshey, A., & Alharbi, A. (2014). A hybrid automatic scoring system for Arabic essays. *AI Communications*, *27*(2), 103–111. https://doi.org/10.3233/AIC-130586

Aljameel, S. S., O'Shea, J. D., Crockett, K. A., & Latham, A. (2017). Survey of string similarity approaches and the challenging faced by the Arabic language. *Proceedings of 2016 11th International Conference on Computer Engineering and Systems, ICCES 2016*, 241–247.

Alsaleem, S. (2011). Automated Arabic Text Categorization Using SVM and NB. In *International Arab Journal of e-Technology* (Vol. 2).

An Automated System for Essay Scoring of Online Exams in Arabic based on Stemming Techniques and Levenshtein Edit Operations. (2016). *International Journal of Computer Science Issues*, *13*(5), 45–50.

Boriah, S., Chandola, V., & Kumar, V. (2008). Similarity Measures for Categorical Data: A Comparative Evaluation. *Proceedings of the 2008 SIAM International Conference on Data Mining*, 243–254.

Contreras, J. O., Hilles, S., & Abubakar, Z. B. (2018). Automated Essay Scoring with Ontology based on Text Mining and NLTK tools. *2018 International Conference on Smart Computing and Electronic Enterprise, ICSCEE 2018*.

Cozma, M., Butnaru, A. M., & Ionescu, R. T. (2018). *Automated essay scoring with string kernels and word embeddings*. Retrieved from http://arxiv.org/abs/1804.07954

Dikli, S., & Russell, M. (2006). An Overview of Automated Scoring of Essays. In *Learning, and Assessment* (Vol. 5). Retrieved from http://www.jtla.org.

Ghandi, M., Supervisor, A., & Kayed, P. A. (2017). *The Semantic Similarity Measures Using Arabic Ontology* .

Gomaa, W. H., & Fahmy, A. A. (2014). Automatic scoring for answers to Arabic test questions. *Computer Speech and Language*, *28*(4), 833–857.

H.Gomaa, W., & A. Fahmy, A. (2013). A Survey of Text Similarity Approaches.

International Journal of Computer Applications, *68*(13), 13–18.

Harika, Y., Sri Latha, I., Sai, V. L., Krishna, P. S., Suneetha, M., Iv, I. /, … Ramakrishna, V. (2017). AUTOMATED ESSAY GRADING USING

FEATURES SELECTION. In *International Research Journal of Engineering and Technology*. Retrieved from www.irjet.net

Higher Education Complex of Bam, & Institute of Electrical and Electronics Engineers. (n.d.). *The 3rd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC2018) : Higher Education Complex of Bam, Bam, Iran.*

Hutchison, D. (2011). Automated Essay Scoring Systems. In *Handbook of Research on New Media Literacy at the K-12 Level* (pp. 777–793). https://doi.org/10.4018/978-1-60566-120-9.ch048

Mohammad, A. H., Alwada‘n, T., & Al-Momani, O. (2016). Arabic Text Categorization Using Support vector machine, Naïve Bayes and Neural Network. *GSTF Journal on Computing (JoC)*, *5*(1). https://doi.org/10.7603/s40601-016-0016-9

Mohler, M., Bunescu, R., & Mihalcea, R. (2011). *Learning to Grade Short Answer Questions using Semantic Similarity Measures and Dependency Graph Alignments*. Retrieved from http://nlp.stanford.edu/software/dependencies

Nath, J., Sanjay, S., & Dwivedi, K. (2013). A Comparative Study on Approaches of Vector Space Model in Information Retrieval. In *International Journal of Computer Applications*.

Osuna, E., Freund, R., & Girosi, F. (2002, November 22). *An improved training algorithm for support vector machines*. 276–285. https://doi.org/10.1109/nnsp.1997.622408

Rababah, H., & Al-Taani, A. T. (2017). An automated scoring approach for Arabic short answers essay questions. *ICIT 2017 - 8th International Conference on Information Technology, Proceedings*, 697–702. https://doi.org/10.1109/ICITECH.2017.8079930

Ramasundaram, S., & Victor, S. P. (2013). Algorithms for text categorization: A comparative study. *World Applied Sciences Journal*, *22*(9), 1232–1240. https://doi.org/10.5829/idosi.wasj.2013.22.09.2670

Rayhan Ahmed, B. M., & Kamal, M. (n.d.). *Mithu (12101115) Syed Mohammad Moinul Islam Robin (13201082)*.

R. M., -, H. H., & -, D. L. (2013). Automated Essay Scoring System By Using Support Vector Machine. *International Journal of Advancements in Computing Technology*, *5*(11), 316–322. https://doi.org/10.4156/ijact.vol5.issue11.37

Schwarm, S. E., & Ostendorf, M. (2007, May 11). *Reading level assessment using support vector machines and statistical language models*. 523–530.

Septem Riza, L., Mulyadi, A., Studi Ilmu Komputer, P., & Pendidikan Matematika dan Ilmu Pengetahuan Alam, F. (n.d.). An Automatic Scoring System for Essay by Using Methods Combination of Term Frequency and n-Gram. In *International Journal of Trend in Research and Development* (Vol. 3). Retrieved from www.ijtrd.com

Shehab, A., Faroun, M., & Rashad, M. (2018). An Automatic Arabic Essay Grading System based on Text Similarity Algorithms. *International Journal of Advanced Computer Science and Applications*, *9*(3).

Srivastava, D. K., & Bhambhu, L. (2005). *DATA CLASSIFICATION USING SUPPORT VECTOR MACHINE*. Retrieved from www.jatit.org

Sundaram, A., Graduate, B. E., Ramnath, R., & Ramanathan, J. (n.d.). *A Comparison of Machine Learning Techniques on Automated Essay Grading and Sentiment*.

Suresh, A., & Jha, M. (2018). Automated Essay Grading using Natural Language Processing and Support Vector Machine. *IJCAT-International Journal of Computing and Technology*, *5*(2). Retrieved from www.IJCAT.org

Surya Vivek Madala, D., Gangal, A., Krishna, S., Goyal, A., Sureka Corresp, A., &

Sureka, A. (2018). *An empirical analysis of machine learning models for automated essay grading An Empirical Analysis of Machine Learning Models for Automated Essay Grading 2*. https://doi.org/10.7287/peerj.preprints.3518v1

Yang, C., Odvody, G. N., Fernandez, C. J., Landivar, J. A., Minzenmayer, R. R., & Nichols, R. L. (2015). Evaluating unsupervised and supervised image classification methods for mapping cotton root rot. *Precision Agriculture*, *16*(2), 201–215. https://doi.org/10.1007/s11119-014-9370-9

Appendix A

Dataset for questions and model answers

id correct_answer Text question_name

'1', ' مجموعه من الاجهزة ', 'عرف الانترنت', 'السؤال الاول'
المتصلة مع بعضها البعض'
'4',هي عمليه ابقاء المعلومات تحت سيطرتك الكامله', 'ماذا يقصد بأمن المعلومات', 'السؤال الثاني''
'3', ' هي برامج متخصصه في الشبكه', 'ماذا نعني بمحركات البحث', 'السؤال الثالث'
الافتراضيه تستخدم للبحث عن المعلومات تساعد الباحث للحصول على المعلومات'
'4', ' جهاز حاسوب , ', 'ما هي متطلبات الاتصال بالانترنت', 'السؤال الرابع'
اشتراك من احد الشركات المزوده للخدمه و متصفح انترنت , موديم'
'5', ' ', 'ماذا يقصد بمزود خدمه الانترنت', 'السؤال الخامس'
شركه تمكن المشترك من الحصول على الانترنت''
'6', ' جهاز ', 'ما هي وظيفه جهاز المودم', 'السؤال السادس'
يربط بين الحاسوب وخط الهاتف'
'7', ' ما هي الخدمات التي تقدمها الحكومه ', 'السؤال السابع'
تتكون الدارة الكهربائية المصباح', 'الالكترونيه , وما فائدتها'
'8', ' ', 'اذكر خدمات البريد الالكتروني', 'السؤال الثامن'
مشاركه الرسائل و الملفات والصور''
'9', 'لان ' ,'السؤال التاسع', 'علل , احترام الحقوق المليكيه الفكريه عند استخدام الانترنت'
حقوق النشر ونسخ المواد الموجوده مملوكه لاشخاص اخرين و لا يحق لاحد ان يعيد نشرها'
'10', ' هي القواعد التي تتحكم', 'عرف بروتوكول الانترنت', 'السؤال العاشر'
في الاتصال و تبادل المعلومات للاجهزه المختلفه المرتبطه بالشبكه'
'11', ' السؤال الحادي عشر', 'عرف الصراع الاجتماعي', '
'عملية اجتماعية تؤدي الى هدم المجتمع'
'12', ' سلك رفيع يوصل في الدارة ', 'عرف المنصهر ', 'السؤال الثاني عشر'
الكهربائية لحماية الاجهزة الكهربائية من الاحتراق وينصهر عندما يمر فيه تيار
قوي'
'13', ' اداه تستخدم لتحديد الاتجاهات وتتركب من', 'عرف البوصلة', 'السؤال الثالث عشر'
مغناطيس صغير يشبه الابرة و يرتكز على سن مدببة تسمح له بالدوران و الاتجاه نحو
الشمال'
'14', ' اجرام سماوية معتمة و ', 'عرف الكوكب', 'السؤال الرابع عشر'
تدور حول الشمس و تستمد ضوءها من الشمس'
'15', ' اذكر مكونات الدارة الكهربائية ', 'السؤال الخامس عشر'
تتكون الدارة الكهربائية المصباح', ''
'16', ' بقايا او اثار ', 'عرف الاحافير ', 'السؤال السادس عشر'
لكائنات حية نباتية او حيوانية عاشت و طمرت في الماضي'
'17', ' كائنات حية قادرة على ', 'عرف الصحالب ', 'السؤال السابع عشر'
تصنيع غذائها بنفسها لاحتوائها على صبغة الكلوروفيل و تعيش في الماء'
'18', ' سرعة ', 'اذكر العوامل المؤثرة في سرعة التبخر', 'السؤال الثامن عشر'
التبخر يتاثر في نوع السائل ودرجة الحرارة وسرعة الهواء المتحرك فوق سطح
السائل'
'19', ' حالات المادة هي ', 'اذكر حالات المادة ', 'السؤال التاسع عشر'
الحالة السائلة والحالة الصلبة والحالة الغازية'

'20', 'عرف البكتيريا', 'السؤال العشرون', ' كائنات حية مجهرية لا نستطبيع
رؤيتها بالعين المجردة وحيدة الخلية و بسيطة التركيب و بدائية النواة'
'21', 'اذكر طرق انتقال الحرارة', 'السؤال الحادي والعشرون ', '
'طرق انتقال الحرارة الحمل و التوصيل و الاشعاع'
'22', ' الاثار', 'ما اثار قبول الاخر على الفرد والمجتمع ', 'السؤال الثاني والعشرون ',
المترتبة على الفرد والمجتمع تحقيق الامن والاستقرار وحفظ الحقوق وعدم التعدي على
الاخرين'
'23', 'ما المقصود بالمواطنة الفاعلة', 'السؤال الثالث والعشرون', ' هي
'مشاركة الافراد في تدبير شؤون مجتمعهم وابداء الراي و القيام بمبادرات هادفة
'24', ' المساواة في', 'اذكر ابرز النتائج القانونية للمساواة', 'السؤال الرابع والعشرون',
الحقوق العامة والمساواة امام القانون و المساواة في الحماية و المساواة في الكرامة
الانسانية'
'25', ' من انماط التغير هي', 'اذكر انماط التغير ', 'السؤال الخامس والعشرون',
التغير على مستوى الفرد والتغير على مستوى الاسرة و التغير على مستوى
المجتمع'
'26', ' من', 'اذكر عوامل التغير الاجتماعي ', 'السؤال السادس و العشرون',
'عوامل التغير الاجتماعي هي العامل السكاني و التقني و البيئي و الاقتصادي
'27', ' التغير في القيم والعادات الاجتماعية', 'ما هي مظاهر التغير الاجتماعي ', 'السؤال السابع و العشرون',
والتغير في الادوار والمراكز الاجتماعية وانتقال من العمل في الرعي و الزراعة الى العمل في القطاعات
الحكومية الخاصة و الصناعية'
'28', ' ما الاسباب الاجتماعية المؤدية الى ظهور مشكلة التغير', 'السؤال الثامن والعشرون',
الاجتماعي ', 'من الاسباب المؤدية الى مشكلة التغير هي ضعف التعليم و عدم قبول الاخر و
مقاومة التغيير'
'29', 'اذكر عوامل التغير في المجتمع الاردني', 'لسؤال التاسع والعشرون', ' عوامل التغير
في المجتمع الاردني هي الهجرات السكانية المتتابعة و التطور التكنولوجي السريع و التطور
في التعليم'
'30', ' من اهم المشكلات', 'اذكر مشكلات التغير الاجتماعي ', 'السؤال الثلاثون ',
التي ظهرت مع التغير الاجتماعي هي المشكلات البيئية و المشكلات و الاسرية و
السلوكية'
'31', ' هي كل الخطط والبرامج', 'عرف السلامه المرورية', 'السؤال الحادي والثلاثون',
المرورية والاجراءات الوقائية المتبعة لتقليل الحوادث المرورية وضعت لحماية الانسان
وممتلكاته وحماية الوطن'
'32', ' من اهم عناصر', 'اذكر عناصر المرور', 'السؤال الثاني والثلاثون',
'المرور العنصر البشري كالمشاة و السائقين و الامنة و الطريق المجهز
'33', ' لوحات معدنية ذات اشكال هندسية متنوعة', 'عرف الشواخص ', 'السؤال الثالث والثلاثون',
توضع على جانبي الطريق على مسافات معينة وارتفاعات محددة ومنها الشواخص التحذيرية
والارشادية'
'34', 'عرف الصراف الالي', 'السؤال الرابع والثلاثون', ' جهاز يستخدم تقنية اتصالات عالية السرعة
و امنة للتعامل بين مقدم الخدمة و متلقي الخدمة للعميل ويستخدم لاجراء المعاملات المالية التي يحتاجها
العميل'
'35', 'السؤال الخامس والثلاثون', 'عرف التسويق السياحي', 'نشاط متكامل يضم جميع
الجهود المبذولة لجذب انتباه المزيد من السائحين المحليين والخارجيين لزيادة الاماكن السياحية
في المملكة'
'36', ' سلسلة من التفاعلات', 'عرف الاحتراق ', 'السؤال السادس والثلاثون',
الكيميائية بين مادتين او اكثر ينتج عنها حرارة و لهب و ضوء و انبعاث
غازات'

'37', ' يتكون', 'اذكر العناصر الذي يتكون منها الحريق ', 'السؤال السابع والثلاثون'
الحريق من عدة عناصر منها مادة قابلة للاشتعال و غاز الاوكسجين و مصدر
للحرارة'
'38', ' هي اسطوانة معدنية', 'عرف طفاية الحريق ', 'السؤال الثامن والثلاثون'
مملوءة بالماء او بمواد كيميائية ثقيلة و عازلة تقوم بعزل الاكسجين عند المادة
المحترقة'
'39', ' هو مخرج لتزويد القوابس الموصولة', 'عرف المقابس ', 'السؤال التاسع والثلاثون'
بالاجهزة الكهربائية بالتيار الكهربائي ويكون موصول من خلال اسلاك التمديدات
الكهربائية'
'40', ' ', 'اذكر سلبيات بطاقة الصراف الالي ', 'السؤال الاربعون '
امكانية الاستخدام من شخص اخر غير صاحبه''

Dataset of Model answers

Id, score, question_id, Answer_Text,

'1', '100', '1', 'مجموعه من الحواسيب ترتبط فيما بينها بواسطه خطوط اتصال لها القدره على مشاركه البيانات'

'2', '75', '1', 'مجموعه من الحواسيب ترتبط مع بعضها بوساطه خطوط اتصال'

'3', '25', '1', 'مجموعه من الحواسيب ترتبط مع بعضها البعض'

'4', '100', '2', 'هي عمليه ابقاء المعلومات تحت سيطرتك الكامله'

'5', '75', '2', 'ابقاء المعلومات تحت سيطره شخص'

'6', '25', '2', 'ابقاء المعلومات بحوزتك ومن دون تدخل'

'7', '100', ' هي برامج متخصصه في الشبكه الافتراضيه تستخدم للبحث عن المعلومات تساعد الباحث للحصول على
3', 'المعلومات'

'8', '75', '3', 'تساعد الباحث للحصول على المعلومات'

'9', '25', '3', 'هي عباره عن برامج بحث في الانترنت'

'10', '100', '4', 'جهاز حاسوب , موديم , اشتراك من احد الشركات المزوده للخدمه و متصفح انترنت'

'11', '75', '4', 'جهاز حاسوب شخصي , او خط هاتف ,و بعض المتطلبات الاخرى'

'12', '25', '4', 'موديم وجهاز'

'13', '100', '5', 'هي الشركه التي توفر لعملائها امكانيه الوصول الى الانترنت عن طريق الاشتراك'

'14', '75', '5', 'شركه تمكن المشترك من الحصول على الانترنت'

'15', '25', '5', 'امكانيه الاشتراك بالانترنت'

'16', '100', ' جهاز يربط بين الحاسوب وخط الهاتف و يقوم بتحويل البيانات الصادره من جهاز الحاسوب الى اشارات
6', 'موجيه تنتقل الى خط الهاتف'

'17', '75', '6', 'هو جهاز يستخدم لربط الحاسوب و خط الهاتف و قد يكون في الداخل او الخارج'

'18', '25', '6', 'جهاز يستخدم لربط الحاسوب و خط الهاتف'

'19', '100', ' تقديم المعاملات الحكوميه و الخدمات العامه عبر الانترنت فائدتها تحسين الاداء الحكومي التقليدي وتقليل الوقت و التكلفه', '7'

'20', '75', '7', 'تقديم الخدمات و المعاملات لتوفير الوقت'

'21', '25', '7', 'تقديم المعاملات و انهائها دون الحاجه الى اللجوء الى الذهاب الى الدائره الحكوميه'

'22', '100', '8', 'مشاركه الرسائل و الملفات والصور'

'23', '75', '8', 'تبادل الرسائل بين مرسل و مستقبل'

'24', '25', '8', 'عمليه تبادل الرسائل بين اكثر من شخص'

'25', '100', '9', 'لان حقوق النشر ونسخ المواد الموجوده مملوكه لاشخاص اخرين و لا يحق لاحد ان يعيد نشرها'

'26', '75', '9', 'لانها ملك لاشخاص ولا يحق التعديل عليها'

'27', '25', '9', 'هذا يعتبر تدخل بخصوصيه الاشخاص'

'28', '100', '10', 'هي القواعد التي تتحكم في الاتصال و تبادل المعلومات للاجهزه المختلفه المرتبطه بالشبكه'

'29', '75', '10', 'هي التحكم بتواصل وتبادل المعلومات'

'30', '25', '10', 'هو التحكم بالحاسوب عند الاتصال بالشبكه'

'31', '100', ' عملية سلبية هدامة لانه يعبر عن القوى الاجتماعية ومدى تصادمها وينشأ نتيجة الضروف الاجنماعية والاقتصادية والسياسية غير المستقرة', '11'

'32', '75', '11', 'عملية اجتماعية تحدث نتيجة الاوضاع السياسية و الاقتصادية و الاجتماعية غير المستقرة'

'33', '25', '11', 'عملية اجتماعية تؤدي الى هدم المجتمع'

'34', '100', ' سلك رفيع يوصل في الدارة الكهربائية لحماية الاجهزة الكهربائية من الاحتراق وينصهر عندما يمر فيه تيار قوي', '12'

'35', '75', '12', 'سلك رفيع لحماية الاجهزة الكهربائية'

'36', '25', '12', 'سلك يوصل في الدارة الكهربائية'

'37', '100', ' اداه تستخدم لتحديد الاتجاهات وتتركب من مغناطيس صغير يشبه الابرة و يرتكز على سن مدببة تسمح له بالدوران و الاتجاه نحو الشمال', '13'

'38', '75', '13', 'اداة تستخدم لتحديد الاتجاهات و تتركب من مغناطيس صغير يشبه الابرة'

'39', '25', '13', 'يستخدم لتحديد الاتجاهات'

'40', '100', '14', 'اجرام سماوية معتمة و تدور حول الشمس و تستمد ضوءها من الشمس'

'41', '75', '14', 'اجرام سماوية معتمة تدور حول الشمس'

'42', '25', '14', 'جرم سماوي معتم'

'43', '100', '15', 'تتكون الدارة الكهربائية من المصباح و السلك و البطارية و المفتاح'

'44', '75', '15', 'تتكون الدارة الكهربائية من المصباح و السلك'

'45', '25', '15', 'تتكون الدارة الكهربائية المصباح '

'46', '100', '16', 'بقايا او اثار لكائنات حية نباتية او حيوانية عاشت و طمرت في الماضي'

'بقايا او اثار لكائنات حية نباتية', '16' ,'75' ,'47'

'اثار لكائنات حية او بقايا', '16' ,'25' ,'48'

'كائنات حية قادرة على تصنيع غذائها بنفسها لاحتوائها على صبغة الكلوروفيل و تعيش في الماء', '17' ,'100' ,'49'

'كائنات حية قادرة على تصنيع غذاءئها بنفسها', '17' ,'75' ,'50'

'كائنات حية تعيش في الماء', '17' ,'25' ,'51'

'سرعة التبخر يتاثر في نوع السائل ودرجة الحرارة وسرعة الهواء المتحرك فوق سطح السائل', '18' ,'100' ,'52'

'سرعة التبخر يتاثر في نوع السائل ودرجة الحرارة', '18' ,'75' ,'53'

'سرعة التبخر يتاثر في نوع السائل', '18' ,'25' ,'54'

'حالات المادة هي الحالة السائلة والحالة الصلبة والحالة الغازية', '19' ,'100' ,'55'

'حالات المادة هي الحالة السائلة والصلبة', '19' ,'75' ,'56'

'حالات المادة هي الحالة السائلة', ' 19' ,'25' ,'57'

كائنات حية مجهرية لا نستطبيع رؤيتها بالعين المجردة وحيدة الخلية و بسيطة التركيب و بدائية النواة', ' ,'100' ,'58'
20''

'كائنات حية بدائية و مجهرية وحيدة الخلية', '20' ,'75' ,'59'

'كائنات حية بدائية', '20' ,'25' ,'60'

'طرق انتقال الحرارة الحمل و التوصيل و الاشعاع', '21' ,'100' ,'61'

'طرق انتقال الحرارة التوصيل و الحمل', '21' ,'75' ,'62'

'طرق انتقال الحرارة الاشعاع', '21' ,'25' ,'63'

الاثار المترتبة على الفرد والمجتمع تحقيق الامن والاستقرار وحفظ الحقوق وعدم التعدي على الاخرين', ' ,'100' ,'64'
22''

'الاثار المترتبة تحقيق الامن و الاستقرار و حفظ الحقوق', '22' ,'75' ,'65'

'الاثار المترتبة تحقيق الامن و الاستقرار و الطمانينة', '22' ,'25' ,'66'

'هي مشاركة الافراد في تدبير شؤون مجتمعهم وابداء الراي و القيام بمبادرات هادفة', '23' ,'100' ,'67'

'هي مشاركة الافراد في تدبير شؤون مجتمعهم وابداء الراي الهادف', '23' ,'75' ,'68'

'هي تدبير شؤون المجتمع', '23' ,'25' ,'69'

المساواة في الحقوق العامة والمساواة امام القانون و المساواة في الحماية و المساواة في الكرامة الانسانية', ' ,'100' ,'70'
24''

'المساواة في الحقوق امام القانون و الحماية و المساواة في الكرامة', '24' ,'75' ,'71'

'المساواة في الحقوق امام القانون', '24' ,'25' ,'72'

من انماط التغير هي التغير على مستوى الفرد والتغير على مستوى الاسرة و التغير على مستوى ' ,'100' ,'73'
'المجتمع', '25

'من انماط التغير هي التغير على مستوى الفرد و التغير على مستوى الاسرة', '25' ,'75' ,'74'

'75', '25', '25', 'من انماط التغير هي التغير على مستوى الفرد'

'76', '100', '26', 'من عوامل التغير الاجتماعي هي العامل السكاني و التقني و البيئي و الاقتصادي'

'77', '75', '26 ', 'من عوامل التغير الاجتماعي هي العامل السكاني و التقني و البيئي'

'78', '25', '26', 'العامل السكاني هو من عوامل التغير الاجتماعي'

'79', '100', ' التغير في القيم والعادات الاجتماعية والتغير في الادوار والمراكز الاجتماعية وانتقال من العمل في الرعي
27', 'و الزراعة الى العمل في القطاعات الحكومية الخاصة و الصناعية'

'80', '75', '27', 'التغير في القيم و العادات الاجتماعية و المراكز الاجتماعية و الانتقال الى القطاع الحكومي'

'81', '25', '27', 'الانتقاال الى العمل في القطاع الحكومي و الخاص و الصناعي'

'82', '100', '28', 'من الاسباب المؤدية الى مشكلة التغير هي ضعف التعليم و عدم قبول الاخر و مقاومة التغيير'

'83', '75', '28', 'من الاسباب الاجتماعية المؤدية الى مشكلة التغير هي ضعف التعليم و التغيير'

'84', '25', '28', 'من الاسباب الاجتماعية المؤدية الى مشكلة التغير هي المقاومة'

'85', '100', ' عوامل التغير في المجتمع الاردني هي الهجرات السكانية المتتابعة و التطور التكنولوجي السريع و
29', 'التطور في التعليم'

'86', '75', '29', 'عوامل التغير في المجتمع الاردني هي تطور التعليم والتكنولوجيا السريعة'

'87', '25', '29', 'من العوامل المؤثرة على المجتمع هي التعليم'

'88', '100', ' من اهم المشكلات التي ظهرت مع التغير الاجتماعي هي المشكلات البيئية و المشكلات و الاسرية و
30', 'السلوكية'

'89', '75', '30', 'المشكلات البيئية والمشكلات الاسرية من المشكلات التي ظهرت في التغير الاجتماعي'

'90', '25', '30', 'ظهرت عدد من المشكلات مع الاسرة'

'91', '100', ' هي كل الخطط والبرامج المرورية والاجراءات الوقائية المتبعة لتقليل الحوادث المرورية وضعت لحماية
31', 'الانسان وممتلكاته وحماية الوطن'

'92', '75', '31', 'الاجراءات لتقليل الحوادث المرورية او منعها لحماية الانسان و حماية الوطن'

'93', '25', '31', 'تقليل الحوادث المرورية ومنعها'

'94', '100', '32', 'من اهم عناصر المرور العنصر البشري كالمشاة و السائقين و الامنة و الطريق المجهز'

'95', '75', '32', 'المشاة والسائقين والعنصر البشري والمركبة تمثل اهم العناصر للمرور'

'96', '25', '32', 'عناصر المرور هو الطريق المجهز'

'97', '100', ' لوحات معدنية ذات اشكال هندسية متنوعة توضع على جانبي الطريق على مسافات معينة وارتفاعات
33', 'محددة ومنها الشواخص التحذيرية والارشادية'

'98', '75', '33', 'لوحات معدنية ذات شكل هندسي توضع في الشارع'

'99', '25', '33', 'هي عبارة عن لوحات توضع على الشوارع'

'100', '100', ' جهاز يستخدم تقنية اتصالات عالية السرعة و امنة للتعامل بين مقدم الخدمة و متلقي الخدمة للعميل
34', 'ويستخدم لاجراء المعاملات المالية التي يحتاجها العميل'

'101', '75', '34', 'جهاز يستخدم لإجراء المعاملات المالية التي يحتاجها العميل'

'102', '25', '34', 'جهاز يستخدم في البنك'

'103', '100', '35', 'نشاط متكامل يضم جميع الجهود المبذولة لجذب انتباه المزيد من السائحين المحليين والخارجيين لزيادة الاماكن السياحية في المملكة'

'104', '75', '35 ', 'نشاط و يضم جذب انتباه المزيد من السائحين لزيادة الاماكن السياحية'

'105', '25', '35', 'نشاط متكامل للسياح'

'106', '100', '36', 'سلسلة من التفاعلات الكيميائية بين مادتين او اكثر ينتج عنها حرارة و لهب و ضوء و انبعاث غازات '

'107', '75', '36', 'سلسلة من التفاعل الكيميائي بين مادتين او اكثر'

'108', '25', '36', 'تفاعل كيميائي'

'109', '100', '37', 'يتكون الحريق من عدة عناصر منها مادة قابلة للاشتعال و غاز الاوكسجين و مصدر للحرارة'

'110', '75', '37', 'يتكون الحريق من عدة عناصر منها مادة قابلة للاشتعال و غاز الاوكسجين'

'111', '25', '37', 'يتكون الحريق من عدة عناصر منها مادة قابلة للاشتعال'

'112', '100', '38', 'هي اسطوانة معدنية مملوءة بالماء او بمواد كيميائية ثقيلة و عازلة تقوم بعزل الاكسجين عند المادة المحترقة'

'113', '75', '38', 'هي اسطوانة معدنية مملوءة بالماء او مواد كيميائية ثقيلة و عازلة'

'114', '25', '38', 'اسطوانة معدنية يوجد فيها ماء'

'115', '100', '39', 'هو مخرج لتزويد القوابس الموصولة بالاجهزة الكهربائية بالتيار الكهربائي ويكون موصول من خلال اسلاك التمديدات الكهربائية'

'116', '75', '39', 'هو مخرج لتزويد القوابس الموصولة بالاجهزة الكهربائية بالتيار الكهربائي'

'117', '25', '39', 'هو تزويد الاجهزة بالتيار'

'118', '100', '40', 'من السلبيات للصراف الالي ان الجهاز قد يتعرض للعطل ويخرج من الخدمة و امكانية استخدام البطاقة من غير صاحبها في حالة فقدانها و حصر البطاقة داخل الصراف ليعيب في الجهاز'

'119', '75', '40', 'من السلبيات للصراف الالي قد يتعرض للعطل و يخرج من الخدمة و امكانية الاستخدام'

'120', '25', '40', 'امكانية الاستخدام من شخص اخر غير صاحبه'

Appendix B

Pseudo code

```
# encoding=utf-8
import …

from math import sqrt
from sklearn.feature_extraction.text import CountVectorizer,
TfidfTransformer

from nltk.corpus import stopwords
from nltk.corpus import wordnet
from nltk.stem.isri import ISRIStemmer
from nltk.tokenize import word_tokenize
import pandas as pd
import numpy as np
from sklearn.linear_model import SGDClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.pipeline import Pipeline

stopwords = set(stopwords.words("arabic"))


def UncommonWords(A, B):
    # count will contain all the word counts
    count = {}

    # insert words of string A to hash
    for word in A:
        count[word] = count.get(word, 0) + 1

    # insert words of string B to hash
    for word in B:
        count[word] = count.get(word, 0) + 1

    # return required list of words
    return [word for word in count if count[word] == 1]


def word(data):
    data = word_tokenize(data)
    filter = []
    IS = ISRIStemmer()
    for w in data:
        if w not in stopwords:
            stem = IS.stem(w)
            filter.append(stem)
    return filter


def word_net(word_list):
    synonyms = []
    for word in word_list:
        try:
            syn = wordnet.synsets(word, lang=('arb'))[0]
            result = [lemma.name() for lemma in syn.lemmas(lang='arb')]
            synonyms.append(result)
        except:
            synonyms.append(word)
    return synonyms


# word then word_net
def similerty(answer_word, test_word):
```

```
    score = 0
    for words in answer_word:
        for word_s in test_word:
            if words == word_s:
                score += 0.5
    return score

from sklearn.metrics.pairwise import linear_kernel, cosine_similarity
from sklearn.feature_extraction.text import TfidfVectorizer



def find_similar(tfidf_matrix, index, top_n=5):
    cosine_similarities = linear_kernel(tfidf_matrix[index:index + 1],
tfidf_matrix).flatten()
    related_docs_indices = [i for i in cosine_similarities.argsort()[::-1]
if i != index]
    return [(index, cosine_similarities[index]) for index in
related_docs_indices][0:top_n]
def data():
    corpus = []
    file = "answer.csv"
    with open(file, "r") as paper:
        corpus.append((paper.read()))
    tf = TfidfVectorizer(analyzer='word', ngram_range=(1, 3), min_df=0,
stop_words=stopwords)
    tfidf_matrix = tf.fit_transform([content for content in corpus])

    return tfidf_matrix



from difflib import SequenceMatcher



import csv
def write(answer ,q_answer ,score):
    # csv file name
    filename = "answer.csv"
    with open(filename,mode='a',newline='',encoding='utf-8') as csvfile:
        writer = csv.writer(csvfile)
        rows ={"answer": answer, "q_answer": q_answer, "score": score}
        writer.writerow(rows.values())
        csvfile.close()
def doc(doc):
    documents = []
    from nltk.stem import WordNetLemmatizer
    stemmer = WordNetLemmatizer()
    for sen in range(0, len(doc)):
        # Remove all the special characters
        document = re.sub(r'\W', ' ', str(doc[sen]))
        # Substituting multiple spaces with single space
        document = re.sub(r'\s+', ' ', document, flags=re.I)
        # Removing prefixed 'b'
        document = re.sub(r'^b\s+', '', document)
        # Converting to Lowercase
        document = document.lower()
        # Lemmatization
        document = document.split()
        document = [stemmer.lemmatize(word) for word in document]
        document = ' '.join(document)
        documents.append(document)
    return documents

def extract_words(sentence):
    ignore_words = ['a']
    words = re.sub("[^\w]", " ",  sentence).split()
```

```
#nltk.word_tokenize(sentence)
    print('words:' + re)
    print('words:' + words)
    words_cleaned = [w.lower() for w in words if w not in stopwords]
    return words_cleaned


def tokenize_sentences(sentences):
    words = []
    for sentence in sentences:
        w = extract_words(sentence)
        words.extend(w)

    words = sorted(list(set(words)))
    return words


def bagofwords(sentence):
    vectorizer = CountVectorizer(stop_words=stopwords)
    return vectorizer.fit_transform(sentence).toarray()
def SVM():
    Corpus = pd.read_csv("answer.csv","ar")
    test_data = Corpus['answer']
    train_data =Corpus['q_answer']
    test =doc(test_data)
    train =doc(train_data)
    vocabulary_train =tokenize_sentences(train)
    vocabulary_test =tokenize_sentences(test)
    vectorizer = CountVectorizer(stop_words=stopwords)
    X_train_counts = vectorizer.fit_transform(train)
    transformer = TfidfTransformer()
    tfidf_matrix_train = transformer.fit_transform(X_train_counts)
    clf = MultinomialNB().fit(tfidf_matrix_train, test)
    text_clf_svm = Pipeline([('vect', CountVectorizer()),
                             ('tfidf', TfidfTransformer()),
                             ('clf-svm', SGDClassifier(loss='hinge',
penalty='l2',
                             alpha = 1e-3, n_iter = 5, random_state =
42)),])
    text_clf = text_clf_svm.fit(train, test)
    predicted_svm = text_clf_svm.predict(train)
    score= np.mean(predicted_svm == test)
    return str(score)


def word2vec(word):
    from collections import Counter
    from math import sqrt

    # count the characters in word
    cw = Counter(word)
    # precomputes a set of the different characters
    sw = set(cw)

    # precomputes the "length" of the word vector
    lw = sqrt(sum(c*c for c in cw.values()))
   # lw = sum(c * c for c in cw.values())/len(cw)
    print('cw:' + '%.2f' % lw)
    # return a tuple
    return cw, sw, lw

def cosdis(v1, v2):
    # which characters are common to the two words?
    common = v1[1].intersection(v2[1])
    # by definition of cosine distance we have
    return sum(v1[0][ch]*v2[0][ch] for ch in common)/v1[2]/v2[2]
   # return sum(v1[0][ch] * v2[0][ch] for ch in common) /(sqrt(sum(c*c for c
in v1))*sqrt(sum(c*c for c in v1)) )
```

```
import nltk, string
from sklearn.feature_extraction.text import TfidfVectorizer



stemmer = nltk.stem.isri.ISRIStemmer()
remove_punctuation_map = dict((ord(char), None) for char in
string.punctuation)

def stem_tokens(tokens):
    return [stemmer.stem(item) for item in tokens]

'''remove punctuation, lowercase, stem'''
def normalize(text):
    return
stem_tokens(nltk.word_tokenize(text.lower().translate(remove_punctuation_map
)))

vectorizer = TfidfVectorizer(tokenizer=normalize, stop_words=stopwords)

def cosine_sim(text1, text2):
    tfidf = vectorizer.fit_transform([text1, text2])
    return ((tfidf * tfidf.T).A)[0,1]
```