A Topical Classification of Quranic Arabic Text

4 authors:

Mohammed N. Al-Kabi
Zarqa University
129 PUBLICATIONS   **1,835** CITATIONS

SEE PROFILE

Belal Mustafa Abuata
Yarmouk University
38 PUBLICATIONS   **207** CITATIONS

SEE PROFILE

Heider Ahmad Wahsheh
King Faisal University
53 PUBLICATIONS   **630** CITATIONS

SEE PROFILE

Izzat Alsmadi
Texas A&M University San Antonio
325 PUBLICATIONS   **2,087** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Barcode Security View project

The future of programmable security controls View project

A Topical Classification of Quranic Arabic Text

Mohammed N. Al-Kabi
Faculty of Sciences & IT
Zarqa University
Zarqa, Jordan
malkabi@zu.edu.jo

Belal M. Abu Ata
CIS Department
IT & CS Faculty
Yarmouk University
Irbid, Jordan
belalabuata@yu.edu.jo

Heider A. Wahsheh
Computer Science Department
College of Computer Science
King Khalid University
Abha, Saudi Arabia
heiderwahsheh@yahoo.com

Izzat M. Alsmadi
Information Systems Department
College of Computer & Information Sciences
Prince Sultan University
Riyadh 11586, P. O. Box 66833, Saudi Arabia
ialsmadi@cis.psu.edu.sa

Abstract—**The automatic information classification is an important tool used today in many aspects of our life. This tool is used in document classification, speech recognition, handwriting recognition, search engines, data mining, question-answering systems, etc. There are many conducted studies in the area of English and Arabic textual document/sentence classification. However, the literature has one or two primitive studies to classify the Holy Quran Ayāts (verses). This study aims to evaluate the effectiveness of four well-known classification algorithms (Decision Tree, *K*-Nearest Neighbor (*K-NN*), Support Vector Machine (*SVM*) and Naïve Bayes (*NB*)) to classify different Quranic Ayāts according to their topics.**

Keywords- Data mining; Feature selection; SVM; Arabic text classification; Quran text classification; Machine learning;

I. INTRODUCTION

Islam as a religion is based mainly on two sources; first source is the holy Quran which represents the verbatim word of God (Allah) and final revelation to humanity. Therefore it is considered a highly esteemed literary source. The second source Islam based on is Sunnah (Prophet Muhammad sayings). Sunnah includes interpretations of the holy Quran, teachings and way of life of the prophet.

Holy Quran is considered the main miracle of Prophet Muhammad. Prophet Moses miracles includes parting of the Red sea, Prophet Jesus miracles includes curing patients and resurrection of the dead. Moses' miracles and Jesus' miracles are witnessed by some people who live at their time and it is a proof of those Witnesses, and holy Quran asserts the occurrence of these miracles. Holy Quran as a miracle of Prophet Muhammad is different from the miracles of his predecessors (Prophet Moses and Prophet Jesus), since it defies the Arabs in their language at all times. So it is an eternal miracle since it characterizes by it linguistic perfection and inimitability, true prophecies, and validation of recent scientific discoveries.

The literal meaning of the Arabic word "Quran" or "Qur an" or "(Al-Qur ān)" is "the recitation". The holy Quran revelation took over 23 years, where some parts of it revealed at Mecca city and other parts at Medina city in Saudi Arabia. Holy Quran consists of 114 Suras (chapters). Each Sura consists of a number of āyāt (verses). These Suras (Chapters) are varies widely in their lengths, where Sura lengths measured in the number of āyāt. The first sura (Al-Fatiha) which consists of seven verses only, while the second sura (Al-Baqarah) consists of 286 verses [1-3]. The total number of Quranic āyāt (verses) is 6,236 [4]. There are 77,430 Arabic words used in holy Quran, and the total number of unique Arabic words is 18,994. Also there are 12,183 stems; 3,382 lemmas and 1,685 roots [5].

Arabic language is used mainly in the Middle East and North Africa by around 300 million people these days. Arabic language is the 6th most spoken language worldwide, and it is one of united nation six official languages. It is number one Semitic language used today followed by Amharic language which is the official language of Federal Democratic Republic of Ethiopia, followed by Hebrew. Arabic language is really rich and deep, and this makes a number of Muslim scholars believe that this reason behind the revelation of original Quran was in Arabic. You can find many Arabic terms which differ in their meanings but their English translation is a single English word. Examples of such Arabic words are "فؤاد" and "قلب" which means heart in English, but in Arabic they have two different meanings. Also The Arabic words "إنس", "إنسان", "إنسي", "ناس" have different meanings in Arabic and they are not synonyms, but the translator has to translate them to human being(s) in English. In Quran you can find palindrome like in Mudathir Sura (chapter) 74, the third āyā (verse) states (Declare the greatness only of your Lord, "رَبَّكَ فَكَبِّرْ"). The palindrome should be applied on non-vocalized Arabic text, in other words palindrome should be applied on Arabic text free from diacritical marks.

Holy Quran is originally written in Arabic without using diacritical marks. Arabic orthography was under development at the time of Qur'anic revelation. The evolution of Arabic alphabets started at the 7th century to end up in 8th century by Al-Khalil ibn Ahmad al-Farahidi. Therefore Arabic writings at the time of Quran revelation were un-dotted and free from hamzas and diacritical marks.

The literature lacks to studies about the automatic topical classification of Quranic Arabic āyāt (verses). This may be

Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences
December 22 – 25, 2013, Madinah, Saudi Arabia

1 - 272

due to difficulty of the Arabic language (Classical and Modern Standard Arabic), and the Quranic Arabic language. The holy Quran uses a spelling that is different from the Arabic spelling used these days in newspapers, schools, university, Arabic literature which is known as Modern Standard Arabic (MSA). Let us consider the following four Arabic words extracted from Al-Fatiha sura (Chapter 1) as shown in http://www.quranexplorer.com/quran/ and http://corpus.quran.com. The Dagger Alif or superscript Alif (Dagger Alif, "ألف خنجرية") which is written as a short vertical stroke on top of an Arabic letter are used in Arabic words presented in table I. Table I exhibits clearly the difference in the spelling of holy Quran and MSA spelling for the same words.

TABLE I. THE SPELLING OF ARABIC WORDS (MSA AND QURAN)

Sura (Chapter)	Ayah (Verse)	English meaning	Modern Standard Arabic (MSA)	Quranic Arabic
1	2	The Worlds	العالمين	ٱلْعَٰلَمِين
1	4	Owner	مالك	مَٰلِك
1	6	The path	الصراط	
1	7	Path	صراط	

The explorer in the world of Quranic Web sites will notice different Arabic spellings which do not change the Arabic word, therefore the meaning and pronunciation is not change. Websites like http://tanzil.net/#1:1, and http://www.holyquran.net/cgi-bin/prepare.pl?ch=1 use different spellings of the four Arabic words extracted from Al-Fatiha sura (Chapter 1) as shown in table II. Those Websites use normal (Alif, "ا") instead of (Dagger Alif, " ألف خنجرية") in the spellings of the four Arabic words extracted from Al-Fatiha sura (Chapter 1).

TABLE II. DIFFERENT SPELLINGS OF 4 ARABIC QURANIC WORDS

Sura (Chapter)	Ayah (Verse)	English meaning	Modern Standard Arabic (MSA)	Quranic Arabic
1	2	The Worlds	العالمين	العَالَمِين
1	4	Owner	مالك	مَالِك
1	6	The path	الصراط	الصِّرَاط
1	7	Path	صراط	صِرَاط

In table II the last two columns present fully spelling matched Arabic words. The last column in table II presents a fully vocalized (vowelised or vowelled) Arabic words, while the fourth column presents non-vocalized Arabic Words as used in MSA.

Although of the challenges of the Arabic language, there are a decent number of studies about the Arabic textual classification. Most of these studies used Modern Standard Arabic (MSA) datasets, such as the studies of [6-11] etc. We found only one primitive study in the literature dedicated to the topical classification of only two Suras (chapters) of the holy Quran [12].

We fail to find a dataset for holy Quran that we can use in our studies. Therefore this and future studies will include building such dataset. This study is based on a sample of textual dataset of holy Quran. This textual dataset has Arabic Quranic āyāt (verses) which identified by Muslim scholars as belonging to one of the three selected topics (classes) (Ignorant of religion, "الجاهلون بالدين"), (Oneness of God, " توحيد الله") and (Penalty of Apostates, "جزاء المرتدين"). This study aims to identify the best classification algorithm to classify textual Arabic Quranic āyāt (verses). Therefore (Decision Tree, *K*-Nearest Neighbor (*K-NN*), Support Vector Machine (*SVM*) and Naïve Bayes (*NB*)) classification algorithms were tested. Preliminary results show that Support Vector Machine (*SVM*) is the best.

The rest of the paper organized as follows: Section 2 presents a summary to a number of studies related to Arabic textual classification and in particular to the classification of the verses of holy Quran āyāt (verses). Section 3 presents the adopted approach. Section 4 will show the results of testing the four classification algorithms under consideration. Section 5 presents our conclusions and future plans to improve this study.

II. RELATED WORK

This section briefly surveys the work on Arabic textual classification. Also we present at that end of this section a summary to only one study conducted on the field of topical classification of holy Quran verses.

Duwairi [8] study is based on a dataset consisting of 1,000 Arabic documents which vary in length and writing style. Those documents are distributed equally into 10 predefined classes. In her study a distance-based classifier is tested using the 1,000 Arabic documents. Arabic stemming is adopted in Duwairi [8] study to reduce feature dimensionality of the vectors of the 1,000 Arabic documents. She concludes that distance-based classifier is effective and robust classifier.

Duwairi et al. [9-10] conducted those two studies in the field of Arabic text categorization. Duwairi et al. [9] studied the effect of root-based stemming and light stemming to reduce feature dimensionality of Arabic documents. They noticed that root-based stemming leads to reduce a number of semantically different Arabic words to the same three consonants (Triliteral) verbs, and leads to flaws in the process of automatic classification. Light stemming aims to

Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences
December 22 – 25, 2013, Madinah, Saudi Arabia

1 - 273

reduce Arabic words to their stems, and that means to remove only prefixes and suffixes from the Arabic words. Therefore in most cases light stemming would not reduce a number of semantically different Arabic words to the same stem. The *K*-Nearest Neighbor (*K-NN*) classifier was used in their study using a dataset consisting of 15,000 Arabic documents which distributed equally on three classes. Duwairi et al. [9] conclude that using light stemming yields to a more accurate classification results relative to the use of Arabic root-based stemming. Duwairi et al. [10] improves their previous study [9] by adding a new feature reduction technique (word clusters) in addition to the two feature reduction techniques (root-based stemming and light stemming) already tested before. They use the same classifier (*K-NN*), dataset and classes used in [9]. Also they conclude that light stemming technique as feature reduction technique leads to a more accurate classification results relative to the other two feature reduction techniques.

Three automatic text classification algorithms (Naive Bayes (*NB*), *K-NN*, and Rocchio) were tested on a dataset consisting of 1,445 Arabic text documents by [11]. These documents of this dataset are not equally distributed on the nine categories used in their study. They conclude that *NB* classifier is the best, followed by *K-NN* and Rocchio respectively.

Associative classification (*AC*) is tested in the study of [7]. They tested three algorithms Multi-class classification based on association rule (*MCAR*), *NB*, and *SVM*. A dataset consisting of 5,121 Arabic documents is partitioned equally into 7 categories. Alwedyan et al. [7] study concludes that *MCAR* is more accurate to automatically classify Arabic documents than the other two classifiers.

A small sample of the Arabic sayings (Hadith/Hadeeth) of Prophet Mohammed (Peace and blessings of Allah be upon him (PBUH)) is used by [13] as an Arabic dataset to find the best method that can be used to classify different Arabic sayings of the Prophet into their appropriate category. Al-Kabi et al. [13] use 80 Sayings and 12 classes. The six methods (*NB*, inner product, Jaccard, cosine, Dice, and Euclidean) were under test. They conclude that *NB* is the best method that can be used to classify different sayings (Hadith/Hadeeth) of Prophet Mohammed (PBUH) into their appropriate class (book).

Hadi et al. [6] use a dataset consisting of 2,244 Arabic documents. These documents are unequally partitioned into 5 Islamic categories. Hadi et al. [6] study aims to evaluate two automatic classification algorithms (Support Vector Machine (*SVM*) and (*NB*)), They conclude that the effectiveness of *SVM* is better than the effectiveness of *NB*.

Al-Kabi et al. [12] study is considered primitive and a limited in the field of the topical classification of the holy Quran āyāt (verses), since it does not include all the 6,236 versus of the holy Quran. It is limited to Fatiha (7 āyāt (verses)) and Yaseen (83 āyāt (verses)). That means that their study is limited to the topical classification of only 90 āyāt (verses) out of only 6,236 āyāt (verses) consisting the whole

Quran, in other words this study covers around 0.014 of the total number of the holy Quran verses. Also the number of topics used was limited to only 15.

III. METHODOLOGY

This study aims to evaluate the effectiveness of a four well-known classification algorithms (Decision Tree, K-Nearest Neighbor (K-NN), Support Vector Machine (*SVM*) and Naïve Bayes (*NB*)) to classify different Quranic āyāt (verses) according to their topics.

Our framework includes the following procedures:

1. Removal of Arabic Diacritics ('حركات').

2. Removal of Quranic Symbols (such as: 'ج','ظ').

3. Using the manual human topical classification of Quranic āyāt (verses) by [14] to train and evaluate the four classifiers under consideration.

Many books classified different Quranic āyāt (verses) written by Muslim scholars. These books are different widely in their sizes. Some of these books consist of a number of volumes, and some consists of few pages. This study is based on the lexicon of Quran topics by [14]. In this lexicon the Quranic āyāt (verses) are classified into the following 14 main categories shown in table III.

TABLE III.   MAIN CATEGORIES OF ARABIC QURANIC ĀYĀT (VERSES)

i	Arabic Main Categories	English Translation
1	الباب الأول: أركان الإسلام	Part I: Pillars of Islam
2	الباب الثاني: الإيمان	Part II: Faith
3	الباب الثالث: العلوم	Part III: Science
4	الباب الرابع: العمل (أس الحياة)	Part IV: Working (MSI life)
5	الباب الخامس: الدعوة إلى الله	Part V: Call to God
6	الباب السادس: الجهاد	PART VI: Jihad
7	الباب السابع: الإنسان والعلاقات الإجتماعية	Chapter VII: Human and social relations
8	الباب الثامن: العلاقات الأخلاقية	Part VIII: Ethical relations
9	الباب التاسع: تنظيم العلاقات المالية	Part IX: Regulation of financial relations
10	الباب العاشر: العلاقات القضائية	Part X: Judicial relations
11	الباب الحادي عشر: العلاقات السياسية	Part XI: Political relations
12	الباب الثاني عشر: القصص القرآني	Chapter XII: Quranic stories
13	الباب الثالث عشر: الديانات السابقة	Chapter XIII: Previous religions
14	الباب الرابع عشر: تنوع الخطاب الإلهي	Chapter XIV: The diversity of divine discourse

Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences
December 22 – 25, 2013, Madinah, Saudi Arabia

1 - 274

Each of these categories is classified into subcategories. Table IV shows the seven subcategories of the first main topic (Pillars of Islam).

i	Arabic Subcategories of Pillars of Islam	English Translation
1.1	الفصل الأول: التوحيد	Chapter I: Unification
1.2	الفصل الثاني: الدين	Chapter II: Religion
1.3	الفصل الثالث: النبوة (محمد صلى الله عليه وسلم)	Chapter III: The prophet (Muhammad, peace be upon him)
1.4	الفصل الرابع: الصلاة	Chapter Four: Prayer
1.5	الفصل الخامس: الصيام	Chapter V: Fasting
1.6	الفصل السادس: الزكاة والصدقات	Chapter VI: Zakat and alms
1.7	الفصل السابع: الحج والعمرة	Chapter VII: Hajj and Umrah

V. EXPERIMENTS

The experimental results for the dataset showed that the Naive Bayes (*NB*) yields more accurate results than the other three classifiers. In order to evaluate Arabic text Quranic āyāt (verses) classification six measures were used. The holy Quran āyāt (verses) are classified into 14 main classes as shown in table III, and each of these is classified into a number of subclasses. Table IV presents seven subclasses of the main class (Pillars of Islam, "أركان الإسلام"). The three selected classes used in this study are (Ignorant of religion, "الجاهلون بالدين"), (Oneness of God, "توحيد الله") and (Penalty of Apostates, "جزاء المرتدين") are subclasses of the 1.1 subclass (Unification, "التوحيد"). In [14] 1,227 āyāt (verses) belongs to the selected three classes. All these classes are subclasses from (Unification, "التوحيد") class which is a subclass of (Pillars of Islam, "أركان الإسلام") main category as shown in figure 1.
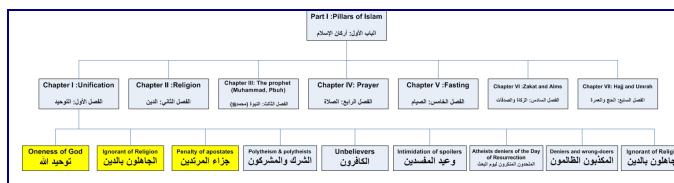

Figure 1. A sample of topical hierarchy of Quranic verses.

In this study, the dataset presents the diversity of the different cases of the distribution of the subclasses for the Quran topics.

The class of (Penalty of Apostates, "جزاء المرتدين"), does not contain any subclasses, where the āyāt (verses) belongs directly to it. While the class of (Ignorant of religion, "الجاهلون بالدين"), is divided into two subclasses (Avoid them,

"الاعراض عنهم"), and (Accept their repentance, "قبول توبتهم"). The class of (Oneness of God, "توحيد الله") contains 20 subclasses. The subclass of (The right to God, "ماندبنا الحق اليه"), contains 6 subclasses, while the subclass of (Divinity, "الألوهية"), contains 5 subclasses, where two of them; (Added Qualities, "الصفات المضافة") contain 81 subclasses, and (Individual Qualities, "الصفات المفردة") which contains 96 subclasses.

The six measurements presented in tables V, VI, VII, and VIII are True Positive (TP) rate, False Positive (FP) rate, Precision (P), Recall (R), F-measure (F1) which combines the previous two measures P and R, Receiver Operating Characteristic (ROC).

Formula 1 shows the Accuracy formula [15].

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad \dots\dots\dots (1)$$

Formula 2 shows the Recall formula [15].

$$Recall_i = \frac{TP}{TP+FN} \quad \dots\dots\dots\dots\dots (2)$$

Formula 3 shows the Precision formula [15].

$$Precision_i = \frac{TP}{TP+FP} \quad \dots\dots\dots\dots\dots (3)$$

Where TP is the number of verses correctly classified as belonging to a class i ("true positive"), FP is the number of verses falsely classified as belonging to a class i ("false positive") and FN is the number of verses falsely classified as not belonging to a class i ("false negative") [15].

The three selected topical classes used in tables V, VI, VII, and VIII are:

Class I: Includes all Quranic āyāt (verses) classified as (Ignorant of religion, "الجاهلون بالدين").
Class II: Includes all Quranic āyāt (verses) classified as (Oneness of God, "توحيد الله").
Class III: Includes all Quranic āyāt (verses) classified as (Penalty of Apostates, "جزاء المرتدين").

Naive Bayes (*NB*) classifier yields an accuracy of 99.9099%, and an error rate of 0.0901% for the dataset used in this study.

Table V shows the values of different measurements which are used to evaluate the effectiveness of Naive Bayes classifier to classify Quranic āyāt (verses) of the dataset into the three selected topical subclasses.

Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences
December 22 – 25, 2013, Madinah, Saudi Arabia

1 - 275

TABLE IV.   EVALUATION RESULTS OF THE NAIVE BAYES CLASSIFIER

Class	*TP*	*FP*	*P*	*R*	*F₁*	*ROC*
I	0.667	0	1	0.667	0.8	0.965
II	1	0.111	1	1	1	0.988
III	0.833	0	0.833	0.833	0.833	0.981
Weighted AVG	0.999	0.111	0.999	0.999	0.999	0.988

Table V shows that the class of (Oneness of God, "توحيد الله") gets the optimal values in the classification. While the class of (Penalty of Apostates, "جزاء المرتدين"), and the class of (Ignorant of religion, "الجاهلون بالدين"), get accepted values for the different measurements.

Table VI shows the six measurements which are used to evaluate the effectiveness of K-Nearest Neighbor (*K-NN*) classifier to classify Quranic āyāt (verses) of the dataset into the three selected topical classes. Using (*K* = 1) yields an accuracy of 99.8198%, and an error rate of 0.1802%.

TABLE V.   EVALUATION RESULTS OF THE *K*-NEAREST NEIGHBOR CLASSIFIER

Class	*TP*	*FP*	*P*	*R*	*F₁*	*ROC*
I	0.667	0	0.677	0.677	0.677	0.785
II	1	0.333	0.999	1	0.999	0.921
III	0.5	0	1	0.5	0.667	0.861
Weighted AVG	0.998	0.332	0.998	0.998	0.988	0.92

Table VI shows that the effectiveness of K-Nearest Neighbor (*K-NN*) classifier is lower than the effectiveness of Naive Bayes. Also Table VI exhibits *K-NN* effectiveness is acceptable.

Table VII shows the six measurements which are used to evaluate the effectiveness of J48 Decision Tree classifier to classify Quranic āyāt (verses) of the dataset into the three selected topical classes. J48 Decision Tree classifier yields an accuracy of 99.5946%, and an error rate of 0.4054% on the dataset used in this study.

TABLE VI.   EVALUATION RESULTS OF THE J48 DECISION TREE CLASSIFIER

Class	*TP*	*FP*	*P*	*R*	*F₁*	*ROC*
I	0	0	0	0	0	0.15
II	1	1	0.966	1	0.998	0.45
III	0	0	0	0	0	0.299
Weighted AVG	0.996	0.966	0.992	0.966	0.994	0.449

Although the decision tree achieved high accuracy results in general, but checking the values in table VII shows that this algorithm is superior in classifying the second class (Oneness of God, "توحيد الله"). Table VII shows that J48 Decision Tree classifier fails completely to classify correctly the other two classes (Penalty of Apostates, "جزاء المرتدين"), and (Ignorant of religion, "الجاهلون بالدين").

The final classifier tested in this study was the Support Vector Machine (*SVM*) classifier, which yields an accuracy of 99.8649%, and an error rate of 0.1351%. Table VIII shows the results of the six measurements adopted in this study to measure the effectiveness of the (*SVM*) classifier to classify Quranic āyāt (verses) belonging to the three selected classes.

TABLE VII.   EVALUATION RESULTS OF THE SUPPORT VECTOR MACHINE (*SVM*) CLASSIFIER

Class	*TP*	*FP*	*P*	*R*	*F₁*	*ROC*
I	0.677	0	0.677	0.677	0.677	0.816
II	1	0.111	1	1	1	0.944
III	0.667	0	0.8	0.667	0.727	0.766
Weighted AVG	0.999	0.111	0.999	0.999	0.999	0.944

Table VIII shows that the Support Vector Machine (*SVM*) classifier yields lower values than the Naive Bayes, but it is more effective than K-Nearest Neighbor (*K-NN*) classifier.

The results of the previous tables show that the Naive Bayes classifier yields the best results to classify the Quranic āyāt (verses) belonging to the three selected classes.

Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences
December 22 – 25, 2013, Madinah, Saudi Arabia

1 - 276

VI. CONCLUSION AND FUTURE WORK

Classification of the Holly Quran is unlike typical classification. First this is since we are dealing with a sacred script and hence classification should be context sensitive. Further, there is a need to understand existed and possible classifications for Quran verses. One of the popular Quran text classification is to classify verses into: (Meccan or Medinan). Other possible classifications that can give deeper understanding of context or semantics should be thoroughly investigated in future research related to the Holly Quran.

In this paper, we measured the effectiveness of four classification techniques ((Decision Tree, (*K-NN*), (*SVM*) and (*NB*)) to classify different Quranic Ayāts or verses according to their topics. We have proposed preprocessing of the classification process by performing the removal of Arabic Diacritics ('حركات') and removal of Quranic Symbols (such as: 'ـً','ج'). We also used the manual human topical classification of Quranic Ayāts by [14] to train and evaluate the four classifiers under consideration. Three selected topics (classes) (Ignorant of religion, "الجاهلون بالدين"), (Oneness of God, "توحيد الله") and (Penalty of Apostates, "جزاء المرتدين") according to [14] are used. This means that 1,227 Ayāts were used in this study out of total number 6236 Ayāts in the whole Quran.

Our evaluation results showed that the *NB* classifier scored the highest accuracy results with lowest error rate whereas the J48 Decision Tree classifier scored the lowest accuracy results with highest error rate. In the future we plan to include all topical classes and subclasses of the holy Quran. That means our future studies will include larger datasets and a larger number of classes and subclasses.

REFERENCES

[1] Quran. Retrieved April 01, 2013, From http://en.wikipedia.org/wiki/Quran

[2] What is Quran? Retrieved April 01, 2013, From http://www.cometoislam.com/quran.htm

[3] What is Quran? Allah describes the Quran in the Quran. Retrieved April 01, 2013, From http://www.iqrasense.com/quran/what-is-quran-allah-describes-the-quran-in-the-quran.html

[4] Islam. Retrieved April 01, 2013, From http://en.wikipedia.org/wiki/Islam

[5] Quran - Wikipedia, the free encyclopedia. Retrieved April 18, 2013, From https://en.wikipedia.org/wiki/Quran

[6] W. Hadi, M. Salam, and J. Al-Widian, "Performance of NB and SVM classifiers in Islamic Arabic data," In Proceedings of the 1st International Conference on Intelligent Semantic Web-Services and Applications (ISWSA '10). ACM, New York, NY, USA, Article 14 , 6, 2010.

[7] J. Alwedyan, W. Hadi, M. Salam, and H. Mansour, "Categorize arabic data sets using multi-class classification based on association rule approach," In Proceedings of the 2011 International Conference on Intelligent Semantic Web-Services and Applications (ISWSA '11). ACM, New York, NY, USA, , Article 18, 8 pages. 2011.

[8] R. Duwairi, "Machine learning for Arabic text categorization," Journal of the American Society for Information Science and Technology, vol. 57, No. 8, pp. 1005–1010, June 2006.

[9] R. Duwairi, M. Al-Refai, and N. Khasawneh, "Stemming Versus Light Stemming as Feature Selection Techniques for Arabic Text Categorization," 4th International Conference on Innovations in Information Technology, IIT '07, pp. 446- 450, 18-20, November 2007.

[10] R. Duwairi, M. Al-Refai, and N. Khasawneh, "Feature reduction techniques for Arabic text categorization," Journal of the American Society for Information Science and Technology, vol. 60, No. 11, pp. 2347–2352, November 2009.

[11] G. Kanaan, R. Al-Shalabi, S. Ghwanmeh, and H. Al-Ma'adeed, "A comparison of text-classification techniques applied to Arabic text," Journal of the American Society for Information Science and Technology, vol. 60, No. 9, pp. 1836–1844, September 2009.

[12] M. AL-Kabi, G. Kanaan, R. Al-Shalabi, K. Nahar, and B. Bani-Ismail, "Statistical Classifier of the Holy Quran Verses (Fatiha & Yaseen Chapters) ," Journal of Applied Sciences, vol. 5, No. 3, pp. 580-583, 2005.

[13] M. Al-Kabi, and S. Al- Sinjilawi, "A Comparative Study of the Efficiency of Different Measures to Classify Arabic Text," University of Sharjah Journal of Pure & Applied Sciences, vol. 4, No. 2, pp. 13 – 24, June 2007.

[14] A. Abu Al-Khair, and M. Kabbani, "Koran teacher intonation," The Tunisian Company for Distribution, 2003.

[15] علي أبو الخير، محمد عربي القباني، مصحف معلم التجويد ، الشركة التونسية للتوزيع، 2003.

[16] I. H. Witten, E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques," Second Edition, Morgan Kaufmann Series in Data Management Systems, 2005.

Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences
December 22 – 25, 2013, Madinah, Saudi Arabia

1 - 277