Keywords Extraction from the Text of Holy Quran Using Linguistic and Heuristic Rules

Fahad Mazaed Alotaibi, Shakeel Ahmad

Faculty of Computing and Information Technology in Rabigh (FCITR) King Abdul Aziz University (KAU) Jeddah
Saudi Arabia

Abstract
The digital age has transformed the way information is extracted from structured and unstructured textual data. Researchers find gaining insights into data for useful and actionable knowledge a great challenge. Natural Language Processing (NLP) is a popular and challenging research area of computer science and artificial intelligence. One key task of NLP is the extraction of keywords from the source text, which is the foundation of a solid search for and understanding of actionable insights. In addition to being used for searching, keywords are used for checking the relevancy of source documents, labeling document clusters, text summarization, text categorization, filtering information, and creating Vector Space Model (VSM) for the training machine learning (ML) model. Search engines also use sophisticated algorithms for searching based on keywords and other optimization parameters. The aim of this research is to extract the keywords from the text of the Holy Quran using linguistic and heuristic rules. The complexity of the NLP task varies across the natural languages. Arabic is a morphologically rich language, in which significant syntactic relations are found at word level. The proposed method is based on possessive (المضاف والمضاف اليه) and genitive (حالة الجر) expressions as well as other important syntactic structures. In the first phase, the model generates unigram, bigram, and trigram expressions, while in the second phase keywords are identified. Keywords are very useful for search engines as well as clustering and generating the summary concepts of documents.

Key words:
Natural Language Processing(NLP), Machine Learning(ML), Vector Space Model(VSM)

1. Introduction

Natural Language Processing (NLP) is a well-researched area of computer science and artificial intelligence. The goal of NLP is to enable the computer to understand the natural language to perform specific target tasks [1]. There are different classes of NLP tasks using lexical analysis, ranging from Named Entity Recognition (NER) to sentiment analysis and machine translation [2]. Over the past two decades, the computational linguistics has grown with unmatched speed and become an exciting research area. Several significant developments have enabled these advances: (a) gigantic computing machines, (b) a large volume of data in linguistics, (c) the power of machine learning methods, and (d) progress in understanding the human languages [3]. The Arabic language is spoken by more than 500 million people worldwide and is one of the six UN official languages [4]; it is also the sacred language of 1.7 billion Muslims [5]. This research is concerned with the Arabic text particularly with the Holy Quran text. Quran is the divine book and ultimate guidance for humanity; it is a book of Islamic principles [6]. The digital age has made it easy for us to store huge amounts of data in machine readable forms [7]. The large number of available options of Arabic textual information demands sophisticated tools and techniques to extract the desired knowledge. The NLP task, known as extraction of keywords, provides the foundation of understanding the source text and gaining actionable insights. It also provides input for other NLP tasks, such as checking the relevancy of documents by measuring keywords similarity, labeling document clusters, text summarization, text categorization, filtering irrelevant information, and creating Vector Space Model (VSM) for the training machine learning (ML) model. Search engines also use sophisticated search algorithms based on keywords and other optimization parameters. These extracted words can be used to search for particular topics in the Quranic text or clustering different verses based on keywords similarity.

Arabic is a morphologically rich language, in which significant syntactic relations are found at word level. The proposed method of keyword extraction is based on the possessive (المضاف) والمضاف اليه and genitive (حالة الجر) expressions and other important syntactic structures of the language. In the Arabic language, a noun (الاسم) has different forms: singular, dual, and plural. Plural nouns and adjectives are further categorized according to gender and into regular and broken ones. The nominative, accusative, and genitive forms of a noun depend on the role of that noun in the sentence [8].

The Holy Quran has more than 77 thousand words arranged in different verses, surah, and chapters [9]. The Quran text for this study was downloaded from the Tanzil Quranic project website [10]. In the first phase, preprocessing tasks such as stop words removal and tokenization as well as unigram, bigram, and trigram expressions are generated, while in the second phase keywords will be identified using different syntactic rules. Figure 1 displays the block diagram of the proposed method.

Manuscript received February 5, 2019
Manuscript revised February 20, 2019

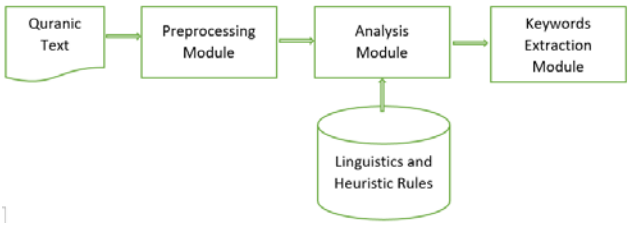

Fig. 1 Keywords Extraction Model

2. Literature Review

Keyword extraction is the process of recognizing meaningful words in a textual document that describes and presents important ideas of the source text. This is one of the NLP tasks that has been studied by many researchers over the past two decades. The key phrase extraction approach presented in [11] is probably the first heuristic attempt. The researchers developed an intelligent agent to extract semantically significant phrases from documents. Four experiments were performed by [12] for "Learning to extract key phrases from the Text." In the first and second experiments, the C4.5 decision tree induction algorithm and the GenEx machine learning-based algorithm were devised by Peter Turney; another system known as key phrase extraction algorithm (KEA) has also been introduced [13]. The goal of KEA is to classify the candidate phrases as key phrases or avoid using the Naïve Bayes learning model. An improved automatic keyword extraction method based on linguistic rules is presented in [14]. This approach is based on supervised learning, first proposed in [13].

The algorithm proposed in [15] can be used for both supervised and unsupervised keyword extraction. A considerable number of NLP researchers have worked on keyword identification in documents related to different natural languages. The work in [16] presents a system called KP-Miner for extracting key phrases from English and Arabic documents. KP-Miner can be configured according to the general nature of documents and key phrases. The method for assigning weight to candidate keywords is presented by Liu et al. [17]. Their weighting scheme is based on the vector of four features: (a) TF-ITF, (b) Part of Speech (POS) tagging, (c) word relative position of first occurrence, and (d) Chi-Square (X2) Statistic. They conducted many experiments and compared the results for both manual and automatic assignments. The automatic method of assignment has illustrated 64% precision, 73% recall, and 68% F-measure. Arabic has gained great attention from the research community worldwide.

Word co-occurrence-based method for the extraction of Arabic keywords is presented in [18]. In this method, the selection of candidate keywords is based on TF-ITF and Chi-Square (X2) Statistic. Ultimately, a candidate word is selected as the keyword, which satisfies the specified threshold and rules. The candidates that have high X2 are considered keywords. The unsupervised approach for extracting keywords from Arabic documents has been proposed in [19]; this is a two-phase approach that combines the statistical analysis and linguistic rules for the domain-independent extraction of keywords in a single document. Mahdaouy et al. [20] have proposed the hybrid method for extracting multi-word terms from Arabic texts. This method is based on two filters; first, the POS tagger is used to extract the candidates, and, second, the syntactic pattern is identified, which is based on a statistical measure called NLC-value. The work in [21] presents a novel method for extracting keywords from Arabic documents using term equivalence classes. This approach combines the multiple levels of statistical analysis with linguistic analysis and provides domain-independent extraction.

The new supervised algorithm, known as Automatic Key phrases Extraction from Arabic (AKEA), is presented by [22]. The experiments in this study were performed on a data set that had been collected from Wikipedia; the AKEA shows high accuracy in recognizing two-word phrases. The ontological model of the related Quranic concepts described in different chapters is presented in [23]. The authors used Protocol and RDF Query Language (SPARQL) queries to extract the related concepts belonging to different sections of Quran. Research [24] proposes a method for extracting keywords from the Arabic text based on the bag-of-concept method. The proposed algorithm uses the semantic vector space model and creates a matrix called word-context to group similar words together in the same class. This paper proposes a method for extracting keywords from the text of the Holy Quran using linguistic and heuristic rules.

3. Design and Methodology

Figure 2 depicts the overall modules of the proposed methodology. This model consists of three main modules, namely text preprocessing, analysis, and keywords extraction and selection.

3.1 Text Preprocessing

This module performs various preprocessing tasks such as removal of stop words, word tokenization, creation of bigram and trigram terms and other preprocessing tasks required for the identification of syntactic structures.

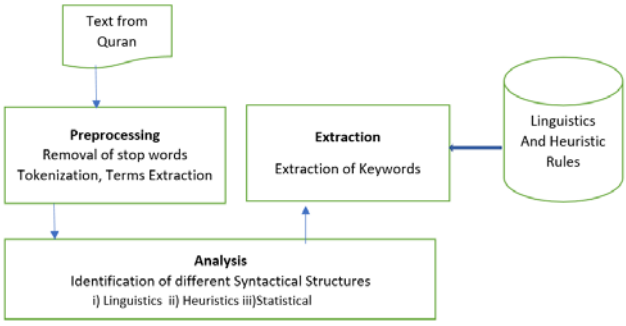

Fig. 2 Proposed Methodology

3.2 Rules for Keywords

In this module, different syntactic structures are identified by applying different linguistic and heuristic rules. Table 1 shows several possessive (المضاف والمضاف اليه) expressions from the Holy Quran, which are strong candidates for keywords.

Table 1: Possessive (المضاف والمضاف اليه) from Holy Quran

المضاف والمضاف اليه	سورة:ايت(Surah:Ayat)
إِذَا جَاءَ نَصْرُ اللَّـهِ	110:1
لَيْلَةُ الْقَدْرِ خَيْرٌ مِّنْ أَلْفِ شَهْرٍ	97:3
أُولَٰئِكَ أَصْحَابُ النَّارِ	2:39
ذَٰلِكَ يَوْمُ الْخُرُوجِ	50:42
هَٰذِهِ نَاقَةُ اللَّـهِ	7:73
مُحَمَّدٌ رَّسُولُ اللَّـهِ	48:29

3.2.1 Based on Genitive Compound

A complete sentence is divided into a bigram, consisting of two words, w1 and w2. Now, the following genitive compound rule is applied to detect the keyword:

$$S = \cup_{x=1}^{n} W_x$$

$$KeyW_{Rule-1} = \cup_{i=1}^{n}\{W_i + W_{i+1}, \quad if\ LCW_i \in \text{Kasrah (ِ)or Dumma (ُ)} and\ LCW_{i+1} \in \text{Kasrah (ِ)} \quad (1)$$

This equation indicates that, in a bigram, if the last character of the first word LCW_i ends with a kasrah or dumma and the last character of the second word ends with a kasrah, then the combination of these words (i.e., $W_i + W_{i+1}$) can be considered the keyword: $KeyW_{Rule-1}$ as Rule-1.

3.2.2 Based on Prepositional Compound

A predefined prepositional compound list in [25] has generated Unicode for the words shown in Table-2.

Table 2: List of Prepositional Compound

Words	Unicode
عَلٰى	'\u0639\u0618\u0644\u0670\u0649'
إِلَى	'\u0627\u0650\u0644\u0670\u0649'
مِنْ	'\u0645\u0650\u0646\u06E1'
عَنْ	'\u0639\u0618\u0646\u06E1'
بِ	'\u0628\u0650'
كَ	'\u0643\u0618'
لِ	'\u0644\u0650'
فِيْ	‘\u0641\u0650\u0649\u06E1'

A bigram can be considered a prepositional compound if the first word is on the prepositional compound list (i.e., عَلٰى جَبَلٍ.) This list also consists of words with single and multiple characters. Thus, a word that starts with a prepositional compound of a single character can also be considered a keyword (i.e., كَرَجُلٍ.) Hence, we formulated two rules by dividing the prepositional compound list into two parts: (a) PC_1, containing "فِيْ مِنْ عَلٰى إِلَى عَنْ" and (b) PC_2, containing "بِ كَ لِ":

$$KeyW_{Rule-2} = \cup_{i=1}^{n}\{W_i + W_{i+1}, \quad if\ W_i \in PC_1 \quad (2)$$

$$KeyW_{Rule-3} = \cup_{i=1}^{n}\{W_i \ , \quad if\ FCW_i \in PC_2 \quad (3)$$

3.2.3 Based on Demonstrative Compound

A predefined demonstrative compound list in [25] has generated Unicode for the words in Table-3.

Table 3: List of Demonstrative Compound List

Words	Unicodes
هٰذَا	'\u06BE\u0670\u0630\u0618\u0627'
هٰذَانِ	'\u06BE\u0670\u0630\u0618\u0627\u0646\u0650'
هٰذَيْنِ	'\u06BE\u0670\u0630\u0618\u06CC\u06E1\u0646\u0650'
هٰذِهِ	'\u06BE\u0670\u0630\u0650\u0665\u0656'
هَاتَانِ	'\u06BE\u0618\u0627\u062A\u0618\u0627\u0646\u0650'
هَاتَيْنِ	'\u06BE\u0618\u0627\u062A\u0618\u06CC\u06E1\u0646\u0650'
ذٰلِكَ	'\u0630\u0670\u0644\u0650\u0643\u0618'
ذٰيْكَ	'\u0630\u0670\u0646\u0650\u0643\u0618'
ذَيْنِكَ	‘\u0630\u0618\u06CC\u06E1\u0646\u0650\u0643\u0618'
تِلْكَ	'\u062A\u0650\u0644\u06E1\u0643\u0618'
تَانِكَ	'\u062A\u0618\u0627\u0646\u0650\u0643\u0618'
هٰؤُلَاءِ	'\u06BE\u0670\u0653\u0648\u0674\u064f\u0644\u0618\u0627\u0653\u0621\u065F'
كِ ألْ	'\u0627\u064f\u0644\u0670\u0653\u0655\u065F\u0643\u0618'

There are two possible rules based on demonstrative compound words. Suppose that demonstrative compound list DC_L contains all the words in the above table. Suppose word W_i belongs to DC_L; then, if the last character of the previous word (i.e., PCW_{i-1}) ends with a kasrah or dumma and the last character of the next word (i.e., NCW_{i+1}) ends with a dumma, then the combination of the words (i.e., $W_{i-1} + W_{i+1}$) can be considered the keyword (i.e., $KeyW_{Rule-4}$) as Rule-4. Take "الْمَدْرَسَةِ هٰذَا ذَكِيٌّ تِلْمِيْذٌ"; here, the word "هٰذَا" belongs to DC_L, and its previous word,

"الْمَدْرَسَةِ," ends with a kasrah, and the next word, "ذَكِيٌ," ends with a dumma; therefore, the combination "الْمَدْرَسَةِذَكِيٌ" can be considered the keyword. If Rule-4 fails, then W_{i+1} can be considered the keyword:

$$KeyW_{Rule-4} = \bigcup_{i=1}^{n} \begin{cases} W_{i-1} + W_{i+1}, & if\ W_i \in DC_L\ and\ PCW_{i-1} \in \text{Kasrah } (\text{ِ}) \text{or Dumma } (\text{ُ}) and\ NCW_{i+1} \in \text{ Dumma } (\text{ُ}) \\ W_{i+1} \end{cases} \quad (4)$$

If the next and previous words of W_i, which belongs to DC_L, do not end according to the said rule, then another rule will be applied to the next two words of W_i.

Table 4: Algorithm for Proposed Method

```
input ayat
input pc1=u'فِيْ مِنْ عَلٰى إِلٰى عَنْ'
input pc2=u'بِ كَ لِ'
input dc=u'تَانِكَ تِلْكَ ذَيْنِكَ ذٰنِكَ هَاتَانِ هٰذِهِ هٰذَيْنِ هٰذَانِ هٰذَا هَاتَيْنِ ذٰلِكَ أُولَئِكَ هٰؤُلَاءِ تَيْنِكَ'
words = ayat.split()
for i in range(len(words)-1):
   if i<len(words)-1:
      s1=words[i]
      j=i+1
      s2=words[j]
      if (s1 in dcw):
         s1=words[i-1]
         if (s1.endswith(u'\u064f') or  s1.endswith(u'\u0650'))
and s2.endswith(u'\u0650'):
            print " Rule-4: Demonstrative Compound " + s1 + "
" +  s2
         else:
            if (s1[0]== u'\u0628' and s1[1]== u'\u0618')  and
(s1[0]== u'\u0643' and s1[1]== u'\u0618') or (s1[0]==
u'\u0644' and s1[1]== u'\u0650') : # كَ لَ
               print " Rule-3: Prepositional Compound " + s1
            else:
               if s1 in pcw1:
                  print " Rule-2: Prepositional Compound "  + s1 +
" " + s2
               else:
                  if (s1.endswith(u'\u064f') or
s1.endswith(u'\u0650')) and s2.endswith(u'\u0650'):
                     print " Rule-1: Genitive Compound " + s1 +
" " +  s2
```

4. Results and Discussions

To detect keywords using the above mentioned rules, we took an Arabic text from the Quranic Surah of Ya Seen (يس), Surah Ar-Rahman (الرحمن), and Surah Al-Waqi'a (الواقعة). The proposed method found 46 keywords from all the three rules in Ya Seen, 15 from Rule-1, 25 from Rule-2, and six from Rule-3 (see Table-5 for Sample).

Table 5: Keywords from Surah Ya Seen

Rules	Keywords
Rule-1: Genitive Compound	وَالْقُرْآنِ الْحَكِيمِ
Rule-1: Genitive Compound	الْعَزِيزِ الرَّحِيمِ
Rule-3: Prepositional Compound	لِتُنذِرَ
Rule-2: Prepositional Compound	فِي أَعْنَاقِهِمْ
Rule-2: Prepositional Compound	مِن بَيْنِ
Rule-1: Genitive Compound	تُنذِرُ مَن
Rule-2: Prepositional Compound	فِي إِمَامٍ
Rule-1: Genitive Compound	إِلَيْهِمُ اثْنَيْنِ
Rule-2: Prepositional Compound	مِن شَيْءٍ
Rule-2: Prepositional Compound	مِنْ أَقْصَى
Rule-3: Prepositional Compound	لِمَنْ

The proposed method found 85 keywords from all the three rules in Ar-Rahman, 71 from Rule-1, 13 from Rule-2, and 1 from Rule-3. Some of the samples are shown in Table-6.

Table 6: Keywords from Sura Ar-Rahman

Rules	Keywords
Rule-1: Genitive Compound	وَالشَّجَرُ يَسْجُدَانِ
Rule-2: Prepositional Compound	فِي الْمِيزَانِ
Rule-3: Prepositional Compound	لِلْأَنَامِ
Rule-1: Genitive Compound	ذَاتُ الْأَكْمَامِ
Rule-1: Genitive Compound	وَالرَّيْحَانُ فَبِأَيِّ
Rule-1: Genitive Compound	فَبِأَيِّ آلَاءِ
Rule-2: Prepositional Compound	مِن صَلْصَالٍ
Rule-2: Prepositional Compound	مِن مَارِجٍ
Rule-2: Prepositional Compound	مِن نَارٍ
Rule-1: Genitive Compound	فَبِأَيِّ آلَاءِ
Rule-1: Genitive Compound	رَبُّ الْمَشْرِقَيْنِ
Rule-1: Genitive Compound	وَرَبُّ الْمَغْرِبَيْنِ
Rule-1: Genitive Compound	الْمَغْرِبَيْنِ فَبِأَيِّ
Rule-1: Genitive Compound	فَبِأَيِّ آلَاءِ
Rule-1: Genitive Compound	الْبَحْرَيْنِ يَلْتَقِيَانِ
Rule-1: Genitive Compound	يَبْغِيَانِ فَبِأَيِّ

The proposed method found 37 keywords from all the three rules in Al-Waqi'a, 21 from Rule-1, 13 from Rule-2, and three from Rule-3. Some of the samples are shown in Table-7.

Table 7: Keywords from Surah Al-Waqi'a

Rules	Keywords
Rule-3: Prepositional Compound	لِوَقْعَتِهَا
Rule-1: Genitive Compound	فَأَصْحَابُ الْمَيْمَنَةِ
Rule-1: Genitive Compound	أَصْحَابُ الْمَيْمَنَةِ
Rule-1: Genitive Compound	وَأَصْحَابُ الْمَشْأَمَةِ
Rule-1: Genitive Compound	أَصْحَابُ الْمَشْأَمَةِ
Rule-2: Prepositional Compound	فِي جَنَّاتِ
Rule-1: Genitive Compound	جَنَّاتِ النَّعِيمِ
Rule-2: Prepositional Compound	مِن مَّعِينٍ
Rule-1: Genitive Compound	كَأَمْثَالِ اللُّؤْلُؤِ
Rule-1: Genitive Compound	اللُّؤْلُؤِ الْمَكْنُونِ
Rule-1: Genitive Compound	وَأَصْحَابُ الْيَمِينِ
Rule-1: Genitive Compound	أَصْحَابُ الْيَمِينِ
Rule-2: Prepositional Compound	فِي سِدْرٍ
Rule-3: Prepositional Compound	لِأَصْحَابِ
Rule-1: Genitive Compound	وَأَصْحَابُ الشِّمَالِ

Here, we take three documents (i.e., Surah Ya Seen (يس), Surah Ar-Rahman (الرحمن), and Surah Al-Waqi'a (الواقعة)) and a query (غَافِلُونَ فَهُمْ آبَاؤُهُمْ أُنْذِرَ مَا قَوْمًا لِتُنْذِرَ الرَّحِيمِ الْعَزِيزِ تَنْزِيلَ يُؤْمِنُونَ لَا فَهُمْ أَكْثَرِهِمْ عَلَىٰ الْقَوْلُ حَقَّ لَقَدْ) to check the LSI scores of the all text, using the extracted keys (Table-5, Table-6, and Table-7) through our proposed work. Table-8 shows the LSI scores of each document with respect to the query. In the second and third columns, there are several scores for the whole text of the documents and the query. In both methods, the scores of DOC-1 and DOC-2 are the highest,

which suggests that they are very close to the query, whereas DOC-3 is very far from the query. Hence, it is clear that we can search for the closest document to the query using only the keywords instead of the whole document. The second method (using keys) of LSI has a better performance than the first method (using whole text), as the following paragraphs illustrate.

Table 8: LSI Score of Documents with Query

Documents	LSI Score using Whole Text of all Documents	LSI Score using only keys of Documents
DOC-1: Surah Ya Seen (يس)	0.999699598183	0.999095783199
DOC-2: Sura Ar-Rahman (الرحمن)	0.993338898347	0.999642026813
DOC-3: Surah Al-Waqi'a (الواقعة)	-0.0271620432789	-0.0315620627622

Table-9 displays the processed words. Using Method-1, LSI processes 752, 355, and 379 words in all the documents, while through Method-2, LSI processes only 47, 85, and 37 words in memory.

Table 9: Processed Words in Memory

Documents	No of Words in Whole Document	No of Extracted KeyWords
Doc-1	752	47
Doc-2	355	85
Doc-3	379	37

Matrix V and Matrix S have the same size in Method-1 and Method-2, while the rest of the matrix have a smaller size in method-2; in other words, the size of Matrix U in Method-1 is 1216609, while it is only 14161 in Method-2 (see Table-10).

Table 10: Sizes of LSI matrices

LSI-Matrix	Size using Whole text of All Documents	Size using only Keys of All Documents
U	1216609	14161
V	9	9
S	3	3
UK	2206	238
Query Matrix	1103	119
Feature Matrix	3309	357

5. Conclusion

The focus of this research was to develop a model for extracting keywords from the Quranic text using linguistic and heuristic rules. Keywords play a significant role in text mining and can be used as an input for other NLP applications. These words play a vital role in text summarization, text categorization, indexing, building VSM and ML, and clustering different verses in Quran based on the similarity measures. The proposed model extracts the different syntactical structures based on the Quranic expressions, such as possessive (المضاف والمضاف اليه) and genitive (حالة الجر) compounds (المركب); therefore, in this regard, it is helpful in teaching the Arabic grammar with Quranic examples.

Acknowledgment

This articles was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah. The authors, therefore, acknowledge with thanks DSR for technical and financial support.

References

[1] Manning, C., Socher, R., Fang, G. G., & Mundra, R. (2017). CS224n: Natural Language Processing with Deep Learning (Lecture Notes).
[2] Nadkarni, Prakash M, Lucila Ohno-Machado, and Wendy W Chapman. "Natural Language Processing: An Introduction." Journal of the American Medical Informatics Association : JAMIA 18.5 (2011): 544–551. PMC. Web. 10 Dec. 2017.
[3] Julia Hirschberg and Christopher D. Manning. Advances in natural language processing, 2015. [doi: 10.1126/science.aaa8685]
[4] https://nlp.stanford.edu/projects/arabic.shtml [Date: 11-12-2017]
[5] https://en.wikipedia.org/wiki/Arabic [Date: 11-12-2017]
[6] N. Robinson, "Discovering the Qur'an", Georgetown University Press, 2002.
[7] Fouzi Harrag. Text mining approach for knowledge extraction in Sahîh Al-Bukhari. Computer in Human Behavior. 2014, pp. 558-566.
[8] Dr Abdur Raheem. Arabic Course: for English-Speeking Students.
[9] Mohammad Alhawarat et. al. Processing the Text of the Holy Quran: a Text Mining Study. International Journal of Advanced Computer Science and Applications,Vol. 6, No. 2, 2015
[10] http://tanzil.net/download [December 2017]
[11] B. Krulwich, C. Burkey. "Learning user information interests through the extraction of semantically significant phrases.", AAAI Spring Sympo- sium on Machine Learning in Information Access, Stanford, CA, 1996.
[12] P.D. Turney. "Learning to extract keyphrases from text.", Technical Report ERB-1057, National Research Council, Institute for Informa- tion Technology, 1999.
[13] I.H. Witten, et al., "KEA: Practical automatic keyphrase extraction.", The Fourth ACM Conference on Digital Libraries, 1999.
[14] A. Hulth. "Improved automatic keyword extraction given more linguistic knowledge.", Conference on Empirical Methods in Natural Language Processing (EMNLP 2003), 2003.

[15] C. Huang, et al., “Keyphrase extraction using semantic networks structure analysis”, The Sixth International Conference on Data Mining (ICDM ‘06), Hong Kong, 2006.
[16] Samhaa R.El-Beltagy, AhmedRafea. “KP-Miner: A keyphrase extraction system for English and Arabic documents.”, Information Systems, vol. 34, pp. 132– 144, 2009.
[17] Liu, W. and Li, W., “To Determine The Weight in a Weighted Sum Method for Domain-Specific Keyword Extraction.”, International Conference on Computer Engineering and Technology, Singapore. Pp. 11-15, 2009.
[18] Mohammed Al-Kabi et al. “Keyword Extraction Based on Word Co-Occurrence Statistical Information for Arabic Text.”, Basic Sci. & Eng. Vol. 22, No. 1, pp. 75- 95, 2013.
[19] Arafat Awajan. “Unsupervised Approach for Automatic Keyword Extraction from Arabic Documents.”, Conference on Computational Linguistics and Speech Processing. pp. 175-184, 2014.
[20] Mahdaouy, Abdelkader El, Saïd EL Ouatik, and Eric Gaussier. "A Study of Association Measures and their Combination for Arabic MWT Extraction." arXiv preprint arXiv:1409.3005, 2014.
[21] Arafat Awajan. “Keyword Extraction from Arabic Documents using Term Equivalence Classes.”, ACM Trans. Asian Low-Resour. Lang. Inf. Process. 14, 2, Article 7 ,2015.
[22] Hassan M. Najadat et al. “Automatic Keyphrase Extractor from Arabic Documents.”, International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016
[23] A.B.M. Shamsuzzaman Sadi et al. “Applying Ontological Modeling on Quranic “Nature” Domain.”, ICICS, 2016.
[24] Suleiman D, Awajan A. “Bag-of-concept based keyword extraction from Arabic documents.”, Information Technology (ICIT), 8th International Conference. pp. 863-869, 2017. IEEE.
[25] Learn Arabic اللغة العربية, URL: http://arabic.speak7.com/arabic_prepositions.htm (Visit On Date: 05-08-2018)