A HYBRID APPROACH FOR INDEXING AND SEARCHING THE HOLY QURAN

Monther Tarawneh[1] and Emad Al-Shawakfa[2]

Department of Computer Information Systems, Yarmouk University, Irbid, Jordan.
monther.t@yu.edu.jo[1], shawakfa@yu.edu.jo[2]

ABSTRACT

Searching and retrieving information from the Holy Quran have recently received special attention by Muslim and Arab researchers. Many applications and websites offered to perform search through the Quran using keywords, semantic search or even both. Arabic; as the language of the Quran, is having a complex structure and thus the searching process of Arabic scripts is rather more difficult than English. In this paper, we propose a hybrid approach; using a combination of syntactic (keyword) and word semantics, to search through the Quran text. The Quran text was pre-processed and represented in a XML format. A system was implemented using Python, where the concept of regular expressions is used to test the approach with a Graphical User Interface used to manipulate user requests. Good and promising results from the approach were obtained.

KEYWORDS

Arabic NLP, IR, Quran, Keyword search, Semantic search.

1. INTRODUCTION

Information has mostly been stored in digital formats, so that it can be accessed from anywhere at any time [1]. In such a world, the information storage and retrieval has become one of the top research areas, particularly over the period of the past few years [2]. Despite a number of research approaches conducted to customize and improve information retrieval techniques, there is still room for improvements in these techniques [3]. The major aim is to make query processing as natural as possible, so that the users can enter a query into the search engine in any language and the system should be able to respond and retrieve the information within the minimum possible delay.

An information retrieval system should be able to serve the users of all levels of expertise; occasional or experts [4]. In return of a user's query, a Search Engine explores all the possible reachable resources [5] considering entered keywords. The high quality retrieval algorithms within a search engine should, however, be able to narrow down the scope of the retrieved documents and return only those documents which are closely related to the query; a problem that becomes more serious when it comes to languages other than English. This is because most of the searching algorithms for retrieving information were developed to serve Latin-based languages such as English. Therefore, applying these algorithms to other languages poses a clear challenge, due to the fundamental underlying differences in the nature of these languages. So, the performance of any algorithm or application is based on the nature and complexity of the language [6], mainly on the characteristics of the language itself.

Islam is one of the fastest growing religions in the world, where Muslims receive their inspirations and beliefs from two major sources: the Quran; the word of the Almighty Allah (God) and the Hadith; the sayings of the prophet Mohammad (peace be upon him). Nowadays,

both Quran and Hadith can be stored digitally and thus offer a better and efficient way to process them using text analytic techniques. These techniques have the potential to summarize, search, analyze and effectively present different concepts from these resources. There is a need for building intelligent tools that could help Quran readers to search for most relevant and effective results related to a given concept.

In this paper, an initial effort was made to introduce and build an approach that can help in processing and understanding the Quran text. This approach can be further used in other works and could be extended to serve other search processes for Islamic knowledge such as searching the Hadith, Fiqh and even through Islamic History.

2. RELATED WORK

There has been some work on Arabic Natural Language Processing related to the field of search and retrieval of Information from the Holy Quran. Most of the implemented Quran information retrieval systems were based on keyword search [7]. Such implemented systems can be further improved and extended to give more close and accurate results, which has formulated the major objective behind this research.

Many websites and/or online search engines or tools were established to deal with searching through the Quran. Some of these were SearchTruth, Allah.pk, the Islamic Search tools, IntoIslam, … and many others.

The SearchTruth system [8], for instance, searches on the basis of keyword substring matching. No exact word matching is performed, but rather searches are performed for a word that is part of any other word in the Quran. For instance, if you search for the word "حم", all verses that contain a word with the substring "حم" will come up as part of the result page such as "رحم", "لحم", , "حم", … etc.

The guided way technology [9], however, offers searching the Quran using different languages. It searches for exact words in the Quran; thus eliminating all other possibilities of word forms, which may lead to less related results. Chatbot [10] tries to find the most considerable words in the query and performs keyword based search to find the relevant verse in the Quran.

The search engine Islamic Search tools [11] performs search using Google web search Engine to find an exact word match (or partial) based on some Islamic related keywords, such as "Quran", "Islam" and other words. Allah.pk [12] is a search engine that looks for an exact keyword match and provides results in both English translation and Arabic. IslamiCity [13] gives results based on the exact string matching in the translation of the Quran. It gives the Arabic verse, its translation and the phonetic verse. IntoIslam [14] is similar to the search engine Islamic Search Tools and searches using the Google API. IntoIslam provides searching in the form of audio, video, translation or text files.

In most of the used approaches to Arabic IR, light stemming is performed [15]-[16]. To search for Arabic words, three methods are usually applied: full-word-based, stem-based and word synonyms-based. The above mentioned tools search on the basis of keywords by different translations using the syntactic approach. However, a semantic search approach is, on the other hand, beyond the syntactic level of matching keywords [17]-[18]. For example, a search for the word "جنة" using the syntactic approach would give only those verses that has the word "جنة" itself in them. However, using the semantic (*or conceptual*) approach, then using the synonyms of the word "جنة" we could find the word to be reflected in other words such as "فردوس", "عدن" or "دار", thus leading to more verses that are related to the concept of "جنة". Unfortunately, the conceptual search approach has not been fully explored and need more attention.

Many of the available hybrid searches today [19] can be used to combine keyword search and ontology-based search. However, such approaches still force the user to choose either keyword-based or ontology-based search without the ability to merge them. A hybrid approach for

extracting information based on keyword and semantic search was proposed by [20], but deals mainly with the English language and uses Resource Description Framework (RDF) for representing data and SPARQL for querying data.

The concept of XML has been introduced widely to be used in fulfilling a semantic search using semantic rules [21]-[22]. Using this concept, all chapters of the Holy Quran could be easily represented in an XML document, thus making it easier to represent, exchange and process data on the web or within any system that supports the usage and manipulation of XML documents. In addition, a structured database can be easily manipulated when stored in XML format. An XML parser is available in most programming languages of today such as the Document Object Module (DOM) parser built in Java.

The concept of regular expressions (regex) was initially introduced in the Unix operating system and then was further adopted to be used in programming languages and is not related to a specific language by itself. So, regex could be found in languages like C++, C#, Python, … etc. to help in searching for text using patterns [23]. By combining XML parsers with regular expressions, we can build a great searching tool, such as that of [24].

With the existence of different approaches to deal with concepts in the Holy Quran, a unified approach was suggested by [27]. In their proposed approach, the authors have started to identify different formats to map different linguistic annotations into SketchEngine format. According to the authors, their approach can be easily reused to unify the alternative concept tagsets; thus making it easier to provide an accessible Quran resource for linguistic research.

A bilingual (Arabic/English) web-based search tool for searching for around 1100 concepts and/or topics of the Holy Quran was introduced by [28]. In this research, the authors used the Mushaf Al Tajweed from which semantic conceptual information were annotated to each verse. To search for any concept, the user navigates through the ontology tree to the requested concept, from which he/she can obtain a list of tagged verses.

A survey on different projects dealing with searching and Information Extraction from the Holy Quran was studied by [29]. In their study, 11 different projects were studied using an 8-criteria comparison approach for projects that were using ontologies as a means of representation and encapsulation of the knowledge in the Holy Quran. According to the authors, an authentic framework, suitable for ontology applications, was used for the purposes of comparison. Different shortcomings of existing projects were mentioned.

A cloud-based integrated Quran portal was introduced in the work of [30]. In their work, the authors have used the Drupal platform technology to build the portal which can be used to search the Qur'an in more than one language. According to the authors, their work can be reused through a set of APIs accessible through the cloud.

A very good survey of existing ontology projects applied to the Holy Quran was made by [31]. In their paper, the authors have surveyed 12 different projects using a 9-criteria comparison process and found out that most of the existing ontologies are incomplete focusing on some specific domains each with no agreement on the type of formats, technologies to use and how to verify their approaches.

3. METHODOLOGY

Following is the methodology that was used to conduct this research, including data collection, pre-processing and the methods used for storage, searching and ranking of the result.

3.1 Data Collection

The text of the Holy Quran was downloaded from [21] in the format of an XML file with fully vocalized text. The file has only attributes for Chapter number, Verse Number and Verse vocalized Text. This website represents a Quranic project that was launched in early 2007 with a mission of producing a standard Quran text that could serve as a reliable source to be used in different Quran related studies.

As part of the research effort, we had to build the full corpora and add to the XML file other attributes and values that refer to the non-vocalized version of the verses, translation of each verse into English, list of synonyms for Quran words as per the context and the explanation (tafsir) of the Quran words and verses. The translation of the meanings of the Quran words was also added from [25], the explanation was added by authors from [26] and based on the word context and Quranic commentaries on each verse from the tafsir in [26] and the list of synonyms was manually added to the attributes of the XML file. To produce the non-vocalized version of the Holy Quran, a simple program was written to strip the verses' words from diacritic symbols. This effort, which took quite some time, has resulted in the final Quran text database that could be used for the purposes of this, as well as other, research on the Holy Quran.

3.2 Regular Expressions

A Regular Expression is a sequence of characters that define a search pattern using some special characters like (^…, […], $, *…) that can be used to help in searching for text. Regular expressions are found in several search engines, search and replace dialogs of several word processors and text editors. Many programming languages provide regular expression capabilities, some are built-in, such as Perl, JavaScript, Ruby, AWK and TCL, while others offers regular expressions via a standard library, like the .NET languages, Java, Python and C++. Using a regular expression to search for a word that starts at the beginning of a string we use the caret (^), where if the search was for a word that is located at the end of a string, we use the dollar sign ($). For instance, to search for any string that starts with the word "hello", we use the pattern: ^hello; which could produce strings like "hello world", "hello there", … etc., where a search like "world$" would give resulting strings like "hello world", "in the world", … etc.

Patterns are constructed to search through the Quran or specific verses. The search is extended to search using the synonyms of the words to reach meaning levels of words; for example, if a query contains the word "جنة", one may get a result that contains synonym words of "جنة" like [ˈفردوس, ˈنعيم, ˈيمين]. Patterns are built based on regular expression rules. An example of a pattern could be given as [20]:

^([a-zA-Z0-9._-])+@([a-zA-Z0-9.-])+\.([a-zA-Z.]){2,5}$

which is mainly used to match with texts that represent email addresses of the form email_address@organization.toplvl

In this research, we have built a set of patterns based on regular expressions to help in simplifying the search process of Quranic text. We used the Python programming language that has a module which provides regular expression matching operations similar to those found in Perl with powerful functions to conduct matching, replacing, filtering and validating of text operations. Having in mind the right to left property of the Arabic language, the following is an example of a Regular Expression Pattern to search for the word "الله" with its all variations in the Quran like "اللهم", "والله", "تالله", … etc.

(ء|ا|و|ف|ل|ب|ت|ي) الله (م|$)

3.3 Holy Quran Data Set

In this research, we have used the concept of XML to represent the different chapters of the Holy Quran. The 114 chapters (Sura) of the Quran are grouped by the tag <SURA> and ended with the tag </SURA>. Each chapter has a number of verses (Aya) which are represented by the tag <AYA> and ended with </AYA>. Chapter tags has a number of attributes: the kind of the chapter; either a value of "مكية" or "مدنية", an index; holding the order number of the Chapter in Quran, the name of the Chapter, the total number of words in each chapter and the part (Juz'o) number. The AYA tag has an attribute "index" which holds the verse number in the Chapter and the Hizb. AYA has subtag(s) named "word", which stores values related to the word such as its English translation, the explanation (Tafseer) of the word and any existing Synonyms; words in the Quran that have the same meaning, to serve for semantic search.

The XML document of the Quran is stored in a file that is read by the Dom parser. DOM parses the file and builds a tree structure of the XML document called Document Object Model (DOM) tree. DOM3 is a standard for creating and manipulating in memory representation of XML content. Figure 1 gives a sample XML representation of Chapter 108 (Al-Kawthar).

Figure 1. XML representation of Chapter 108 (*Al-Kawther*).

The size of the Quran text is fixed and hence, the complete tree of the Quran is built from the XML document representation as shown in Figure 1. A user may be interested in finding additional relevant information from the search; this includes details about the Chapter, Part, page number, explanation and any other related information; something that can be easily determined from the built tree.

3.4 Searching the Quranic Text

The Quran in its original form has the text fully vocalized; with diacritics on different characters. To search through a fully vocalized text is a tedious process and thus, a non-vocalized version was inserted within the XML document. To search for a given string in the Holy Quran, a system was built supported with a graphical user interface with two inputs: the string to search for and the search region style, which can be either the whole Quran text or a specific Chapter.

A Tokenization process is performed on the string being searched and then stored into an array of Tokens. If the "Chapter" search style was chosen, the given Chapter is then selected and the search is performed in that specific chapter; otherwise, the search will cover all chapters of the Holy Quran. Initially, the search starts by storing the synonyms for every word in the Quran into an array of synonyms. Then, using regular expressions, a word match of every word in the query with the synonyms array for every verse is performed. If a match is found, the index of the verse in a search index array is stored. Once the search is done, the index array is sorted and

the term weight representing the count, or the number of times each index in the array has occurred, is calculated. The verse with the highest term weight is chosen as the most relevant verse to the query. The main purpose behind searching a word by word is to consider the different sequences of words and then apply words conceptual search to improve the efficiency of the search. The algorithm used for our approach to build Al-Baheth Searcher is given in Figure 2.

```
1.  Input the user query to search for.
2.  Perform Tokenization of query words.
3.  Check for the chosen search region,
          If (Chapter Search) then

              Get synonyms of Chapter words from the XML Document

          Else

              Get Synonyms of the Whole Quran from the XML Document

          End if

4.  For each Token in the query perform the following:
      a.  Word match between the token and the Synonym array.
      b.  If a match is found then
              Store the chapter and verse indexes of the matching synonym.

5.  Once search is done, sort the found index array
6.  Perform a match of the chapter index with the top sort from the Quran document.
7.  Display the Verse that matches the index of the searched Token.
```
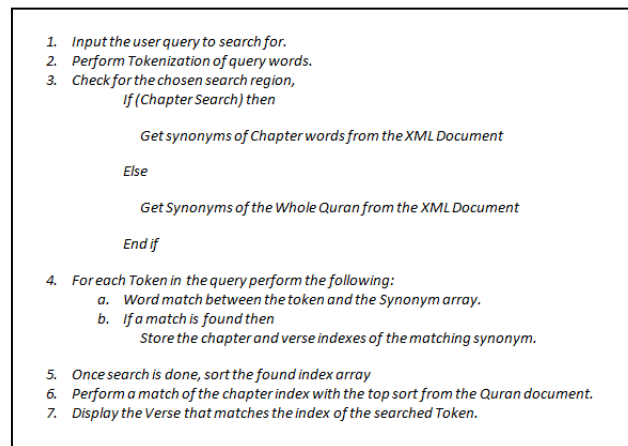
Figure 2. Algorithm used for searching the Quran in Al-Baheth.

Of course, before we could do any search, the Holy Quran text was pre-processed to build the XML document, and this was performed only once. The pre-processing involved removing diacritic symbols to obtain the non-vocalized version of the Quran and applying light stemming on the Quran words to obtain different word forms as part of the Synonyms. Additional synonyms were added to fill the different values of the XML document attributes. While Figure 3 shows the logical flow of the methodology, Figure 4 shows the GUI interface of the built system.
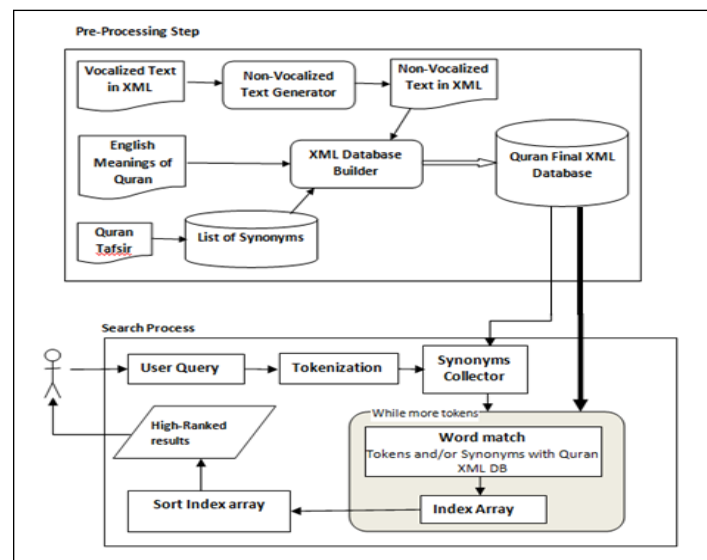

Figure 3. Logical flow of methodology.

47

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 1, No. 1, December 2015.

الباحـث الـقـرأنـي

نطاق البحث
◉ سورة واحدة   النصر
○ القرآن كامل

أدخل النص المراد البحث عنه
الله

نتائج البحث

نص الاية	سورة	رقم الاية	حزب	جزء
إِذَا جَاءَ نَصْرُ اللهِ وَالْفَتْحُ	النصر	1	60	30
وَرَأَيْتَ النَّاسَ يَدْخُلُونَ فِي دِينِ اللهِ أَفْوَاجًا	النصر	2	60	30

Figure 4. Graphical user interface of Al-Baheth searcher of Quran text.

4. EXPERIMENTAL RESULTS

In this research, and for comparison purposes, we have conducted a thorough search with different systems that are known to be used to retrieve text from the Holy Quran and compared their results with the results obtained from our built system; Al-Baheth. Table 1 gives such comparison.

Looking at the results obtained by our approach, in comparison with others, we can say that our approach resulted in obtaining excellent results. For instance, when retrieving the verses that contain the word "الله", the total number of occurrences in the Holy Quran is 2707 and Al-Baheth found 2697 of them with a recall accuracy of 99.6%. Similarly, a recall accuracy of 100% was obtained for the word "الشهر", a 93.2% for the word "الملائكة", a 89.7% for the word "الشياطين", a 98.3% for the word "الدنيا" and a 96.5% recall accuracy for the word "الآخرة".

Table 1. A Comparison of retrieved words between Al-Baheth and other systems.

Al-Baheth	IntoIslam	برنامج القران الكريم جامعة الملك سعود	إحصاء القران الكريم، الإصدار 3	SearchTruth	Guidedways	Allah.Pk	العدد الفعلي	
2697	0	1745	2557	1746	6212	0	2707	الله
12	0	5	6	7	6212	0	12	الشهر
82	0	53	0	53	6212	0	88	الملائكة
79	0	14	0	17	6212	0	88	الشياطين
113	0	111	115	111	6212	0	115	الدنيا
111	0	110	112	115	6212	0	115	الآخرة

From the above results, we can notice that both Allah.pk and IntoIslam did not retrieve any words. The main reason for that is the fact that they are not suitable for searching the Holy Quran in Arabic. Furthermore, "Guidedways" is not suitable as well to search the Holy Quran in Arabic, and that is why the total number of verses in the Quran instead was retrieved as the result. However, when performing search on Guidedways in English, it has outperformed Allah.pk engine which has the capability to search in English.

When searching the Holy Quran using Al-Baheth, both semantic and syntactic information of the words in the query are being used for the purposes of the search using the concept of regular expression patterns. This approach has enhanced the probability of finding our answers via

searching the non-vocalized attribute of each verse combined with the list of synonyms that might be attached to the non-vocalized version of the words of the verse, which has resulted in obtaining better results than other compared approaches.

5. CONCLUSION AND FUTURE RESEARCH

Despite the number of research approaches that have been taken to customize and improve the information retrieval techniques, especially when dealing with Arabic language, there is still room for improvement to these techniques. In this research, we proposed and implemented a hybrid approach; a combination of syntactic- and semantic-based approach to index and search through the Holy Quran. The proposed algorithm efficiently identifies the required part from the Holy Quran that best matches a user's query. We used XML to index the Holy Quran text in a specific structure, which has proven to be an efficient way of representation. The concept of regular expressions was helpful in identifying the requested words because of the flexibility it offers. Results of this work have shown the efficient performance of the proposed algorithm in the retrieval of information from the Quran text. According to the authors' knowledge, no work at this level has been completed for the Arabic language, which could provide the user with the searching capabilities at the level of words. This could refer our approach as the first of its kind on Arabic.

Due to the obtained good results of this approach, and as a future research, it would be helpful to use the approach to search for other topics and concepts in different parts of the Holy Quran. Using this approach, we are planning to conduct a future research for extracting information from different sacred documents like Hadith, Fiqh and even other Islamic historical documents.

REFERENCES

[1] O. Alonso, J. Strötgen, R. A. Baeza-Yates and M. Gertz, "Temporal Information Retrieval: Challenges and Opportunities," TWAW 11, pp. 1-8, 2011.

[2] N. Mitton, S. Papavassiliou, A. Puliafito and K. S. Trivedi, "Combining Cloud and Sensors in a Smart City Environment," EURASIP Journal on Wireless Communications and Networking, pp. 1-10, 2012.

[3] W. Hu, N. Xie, L. Li, X. Zeng and S. Maybank, "A Survey on Visual Content-Based Video Indexing and Retrieval Systems," IEEE Transactions on Man and Cybernetics, Part C: Applications and Reviews, vol. 41, no. 6, pp. 797-819, 2011.

[4] Smirnova and K. Balog, "A User-Oriented Model for Expert Finding," Advances in Information Retrieval, Springer, pp. 580-592, 2011.

[5] T. Sari and A. Kefali, "A Search Engine for Arabic Documents," Colloque International Francophone sur l'Ecrit et le Document, Groupe de Recherche en Communication Ecrite, France, Oct. 2008.

[6] A. El Salam Al Hajjar, M. Hajjar and K. Zreik, "A System for Evaluation of Arabic Root Extraction Methods," Proc. on IEEE 5th International Conference on Internet and Web Applications and Services (ICIW), pp. 506 - 512, Barcelona, Spain, 9-15 May 2010.

[7] S. A. Raza, M. Rehan, A. Farooq, S. M. Ahsan and M. S. Khan, "An Essential Framework for Concept-Based Evolutionary Quranic Search Engine (CEQSE)," Science International, vol. 26, no. 1, 2014.

[8] Search Truth, "Online Quran and Hadith Search Web Portal," last accessed September 1st, 2015 from: www.searchtruth.com.

[9] Guided Ways Technologies, "Online Quran and Hadith Search Web Portal," last accessed July 30th, 2015 from: http://www.guidedways.com/search.php.

[10] B. Abu Shawar and E. Atwell, "An Arabic Chatbot Giving Answers from the Qur'an," Proc. TALN04: XI Conference sur le Traitement Automatique des Langues Naturelles, vol. 2, pp. 197-202, 2004.

[11] IslamicSearch.com, "Islamic Search Powered by Google," last accessed July 12th, 2015 from: http://IslamicSearch.com.

[12] Allah.pk, "The Multilingual Quran and Hadith Search Engine," last accessed August 25th, 2015 from: http://Allah.pk.

[13] IslamiCity.com, "Islam and the Global Muslim eCommunity," last accessed on August 15th, 2015 from: www.quran4theworld.com.

[14] IntoIslam.com, "Islamic Search Engine," last accessed June 10th, 2015 from: http://IntoIslam.com.

[15] K. Darwish and D. W. Oard, "CLIR Experiments at Maryland for TREC-2002: Evidence Combination for Arabic-English Retrieval," DTIC Document, issue LAMP-TR-101, Maryland Univ. College Park Inst. for Advanced Computer Studies, Feb. 2003.

[16] L. S. Larkey and M. E. Connell, "Arabic Information Retrieval at UMass in TREC-10," DTIC Document, Massachusetts Univ. Amherst Center for Intelligent Information Retrieval, 2006.

[17] C. Rocha, D. Schwabe and M. P. Aragao, "A Hybrid Approach for Searching in the Semantic Web," Proc. of the 13th International Conference on World Wide Web, pp. 374-383, New York, NY, USA, 17 – 22 May 2004.

[18] R. Bhagdev, S. Chapman, F. Ciravegna, V. Lanfranchi and D. Petrelli, "Hybrid Search: Effectively Combining Keywords and Semantic Searches," Proc. of the ESWC, Springer, Berlin, Heidelberg, pp. 554-568, 2008.

[19] A. Kiryakov, B. Popov, I. Terziev, D. Manov and D. Ognyanoff, "Semantic Annotation, Indexing, and Retrieval," Web Semantics: Science, Services and Agents on the World Wide Web, vol. 2, no. 1, 2011.

[20] F. Alkhateeb, A. Alzubi, I. A. Doush, S. Aljawarneh and Al E. Maghayreh, "Extracting Authoring Information Based on Keywords and Semantic Search," Proc. of the 1st International Conference on Intelligent Semantic Web-Services and Applications, pp. 1-6, 2010.

[21] Y. Kotb, K. Gondow and T. Katayama, "The SLXS Specification Language for Describing Consistency of XML Documents," Proc. of the 4th International Workshop on Information Computer Science (WICS2002), IEEE Computer Society, 17-18 March 2002.

[22] Y. Kotb, K. Gondow and T. Katayama, "The XML Semantics Checker Model," in: Proc. of the 3rd International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'02), 2002.

[23] elebda3.net, "Regular Expression e-book," downloaded from: http://files.books.elebda3.net/elebda3.net-5647.pdf, on September 2nd, 2015.

[24] A. Arara, A. Smeda and I. Ellabib, "Searching and Analyzing Arabic Text Using Regular Expressions: e–Quran Case Study," International Journal of Computer Science and Electronics Engineering, vol. 1, no. 5, pp. 627-631, 2013.

[25] QuranComplex, "Al-Madinah Mushaf Website," last accessed September 25th, 2015 from: http://publications-img.qurancomplex.gov.sa/?p=44.

[26] Altafsir.com, "Tafsir Website," last accessed on September 28th, 2015 from: http://www.altafsir.com.

[27] N. Abbas, L. Aldhubai, H. Al-Khalifa, Z. Alqassem, E. Atwell, K. Dukes, M. Sawalha and A. Sharaf, "Unifying Linguistic Annotations and Ontologies for the Arabic Quran," in: Proc. of WACL'2 2nd Workshop on Arabic Corpus Linguistics, Lancaster University, UK, 22 July 2013.

"A Hybrid Approach for Indexing and Searching the Holy Quran", Monther Tarawneh and Emad Al-Shawakfa.

[28] N. Abbas and E. Atwell, "Annotating the Arabic Quran with a Classical Semantic Ontology," in: Proc. of WACL'2 2nd Workshop on Arabic Corpus Linguistics, Lancaster University, UK, 22 July 2013.

[29] O. Ahmad, I. Hyder, R. Iqbal, M. Murad, A. Mustapha, N. Sharef and M. Mansoor, "A Survey of Searching and Information Extraction on a Classical Text Using Ontology-Based Semantics Modelling: A Case of Quran," Life Science Journal, vol. 10, no. 4, pp 1370-1377, 2013.

[30] Z. Adhoni and H. Al Hamad, "A Cloud Qur'an Application Using Drupal Technology," International Journal of Web Applications, vol. 6, no. 1, pp 23-38, 2014.

[31] S. M. Alrehaili and E. Atwell, "Computational Ontologies for Semantic Tagging of the Quran: A Survey of Past Approaches," in: Proc. of the 9th International Conference on Language Resources and Evaluations (LREC'14), Reykjavik, Iceland, 26-31 May 2014.

ملخص البحث:

إستحوذت عملية البحث في نصوص القرآن الكريم واسترجاع المعلومات منه على إهتمام العديد من الباحثين العرب والمسلمين في الآونة الأخيرة. و تزخر شبكة الإنترنت بالكثير من التطبيقات والمواقع الإلكترونية التي تفسح المجال للباحث ليقوم بالبحث في القرآن الكريم؛ إما باستخدام الكلمات المفتاحية أو عن طريق معاني الكلمات، أو بالطريقتين معاً.

و الجدير بالذكر أن اللغة العربية، لغة القرآن الكريم، تتميز بتراكيب معقدة تجعل البحث في النصوص العربية أصعب من البحث في النصوص الإنجليزية.

في هذا العمل البحثي، تم اقتراح منحى هجين للبحث في القرآن الكريم يجمع بين البحث باستخدام الكلمات المفتاحية والبحث باستخدام معاني الكلمات. ولتحقيق هذه الغاية، تمت معالجة نصوص القرآن الكريم مسبقاً و تجهيزها بنسيق XML لتكون ملائمة لعملية البحث. و للتحقق من صلاحية المنحى المقترح و ملائمته لعملية البحث و استرجاع المعلومات من القرآن الكريم، تمّ بناء نظام باستخدام لغة بايثون؛ إذ يستخدم مفهوم التعابير المنتظمة لاختبار المنحى المقترح مع توفير واجهة برمجية لمعالجة طلبات المستخدمين. و قد تم الحصول من المنحى المقترح على نتائج جيّدة و واعدة.