International Journal of Recent Research in Social Sciences and Humanities (IJRRSSH)
Vol. 1, Issue 1, pp: (47-53), Month: April - June 2014, Available at: www.paperpublications.org

Adopting Quadrilateral Arabic Roots in Search Engine of E-library System

Said Mohammed Al-Rashdi[1], Dr.S.Arockiasamy[2]

[1] Research Scholar, Information Systems, University of Nizwa, Sultanate of Oman

[2] Head, Information Systems, University of Nizwa

Abstract: **Information retrieval is the method to retrieve information according to user needs. E-library is one of the interesting ways for study and education because it includes a huge amount of information and it is stored in special database or extracting from a corpus of documents. The E-library is a part of an information retrieval system. It provides methods to get information and increase knowledge. But there are inadequacies in the Arabic terms search library and they can be solved by enhancing or adopting algorithm in order to make the search of Arabic language more efficient and easier. In this work, an algorithm for quadrilaterals of Arabic words for use with a search engine of the E-library has been adopted and integrated with New Approach for Extracting Quadrilateral Arabic Root and Pattern – based Stemmer for Finding Arabic Roots. According to the analysis done between ordinary search and quad search, it can observe the Confidence Interval of the Difference (95%) in the ordinal search located between 1.34 and 1.66. In contrast, it located between 1.45 and 1.95 in quad search. So, the result of using algorithm for quadrilateral of the Arabic word is more effective than ordinary search regarding the results in analysis tests.**

1.0 Introduction

Information Systems is a very important field that provides the knowledge for the humanity by using all processes with the body of data. So, it includes many of the research areas supported with business, economics and information technology in order to provide the benefits needed. The physical method of any information systems(IS) is to find the relevant communication between people and IT in order to provide the standard way to maximize the benefits of any application. IS helps the specialist to design and analyze the system and compare with others in order to show the differentiation of the old and latest application. Information retrieval is the method to find the relevant tasks according to the users need. The best example of these methods is search engines because they use this approach to customize the user interest through special algorithms supported with web mining and use them widely over the internet. The improvement of the E-library systems is to increase the number of users and make it more powerful and efficient.

The Arabic language is the fifth most widely spoken language in the word (Kanaan, AL-shalabi& AL-Kabi, 2005) so, it includes many fields to share with technology and it has many different from the English Language.

Arabic Language is strongly connected with the Qur'an which is the holy book of around two billion Muslims around the world. The Arabic language contains twenty-eight letters. It includes many types of morphology that start reading or writing from the right to left and the words includes a connected string of letters according to its position. So, it needs to implement specialized way to establish the relevance among the Arabic texts.

The challenge of that is to make a perfect search for any Arabic words and find the most precise and relevant query result. Also, the complexity is to process the Arabic word because most of Arabic letters written in different forms depending on the position in the word. (Ex: عـ - ـعـ - ـع – ع) .

Table 1.1: The Arabic Letter Division

The Arabic letter	Separate	First write	Middle	Last
Ain(ع)	ع	عـ	ـعـ	ـع

According to the position of the letter it changes the pattern of the word style and this makes it different to identify the root in order to perform a search and make any required to match the user need. Also, there are some complex rules in the use of morphology which can be for triple or quad words.

2.0 Literature Review

2.1 New Approach for Extracting Quadriliteral Arabic Root

According to GhassanKanaan, Riyad Al-Shalabi, Mohammed Naji Al-Kabi compared between many previous studies and showed the differentiation between analyzes of the Arabic methods and structure which assisted me in getting that the roots of the words.

Extracting the quadrilateral of the Arabic words is the main idea in this research as well as the comparison between the factors of the performance for derivatives of the Arabic morphological system. The algorithm used in this publication was tested on 145 Arabic quadrilateral words for verb, and it got 95% measures for the results.

The semantic method is the part of this language that makes Arabic from other computing languages like English. Most Arabic words are trilateral because about 64% of all Arabic words contain three origin letters. The form of templates "أوزان" is the structure used to get the relevant word by using morphology to match the root of any words given.

The model of trilateral word is called "فعل" (faal), and each word with trilateral letter will match the model and get theresult. There are many terms in Arabic that carry no meaning. They are used to join with other words to get correct phrases within Arabic structure such as: stop words, explanations, prepositions, conjunctions, interjections, and exceptions, negatives. The applications to be used for Arabic language need to work under natural language process, computerized language translation, compression of data and spell checking in information retrieval. So, the algorithm needs the rules to show the pattern and root files required large amount of storage space and processing time.

2.2 Pattern – based Stemmer for Finding Arabic Roots

RiyadAlshalabi in his research describes how the Arabic language is very complex and contains many rules. Therefore many methods are needed to extract the root of the Arabic words and to process the item using a search engine or retrieval systems.

This algorithm explains the procedure of the extracted in the trilateral root for Arabic Language and eliminated unnecessary letters including the suffix, prefix and infix.

Pattern-based algorithm was tested on a corpus of 72 abstract words and the accuracy of this algorithm was about 92%.

The algorithm covered the difference between the kinds of the stemming of an Arabic words contain trilateral letters and used light stemmers, morphological analysis and statistical stemmers to reduce the complexity of the implementation and to manually construct dictionary.

Normalization of the corpus was used to convert the specific Arabic letters " إ، أ، ا". For Example: " ا ". So, that this process made performance of the search efficient and also check the vowels, duplicate punctuation and stop words. As mentioned above the Arabic words need to be normalized and steps made to reduce the risk of the problem in the algorithm because the algorithm checks the statements in the algorithm descriptions and extracts the root of the given word.

3.0 Algorithm Description and Methodology

The methodology of the work is to find an efficient way to retrieve the root of quadrilateral Arabic words and show the result in the E-Library System in order to keep the Arabic language closer with the user's needs.

The example of the search methods in the Arabic words is "الهندسة" means "engineering" with prefix "ال" and suffix " ة " and the result of search find 94 results but when we search by using the algorithm of the root and morphology of the same word "هندس" the result is about 163. The result is so far the ordinary word compare with the morphology or root of the same word.

The algorithm, that is adopting the E-library system uses to compare between the ordinary word and the root of the same word and the result of the two comparison extract the differentiation of the previous compare.

PseudoCode

The structure of the algorithm uses array to remove the prefix and suffix from the query of the word and the code shown below:

```
//root search
echo"<br><input type=checkbox name=root value=1";if($root)echo" checked";echo">البحث بأصل الكلمة الرباعي";
function wordRoot($word){
    $word=trim($word);
    $prefix=array("كال","بال","الم","ال","لل","يت","ت","ا","أ","إ","آ","م");
    $suffix=array("ها","ون","تي","ين","ان","اة","ات","ة","ه","ي","ى","ا","أ","إ","آ");
    $exception=array("","انحدر","عمان","دارس","الدارسين","الله");
    $exception4=array("","تلتل","مركش");

    if(strlen($word)<4)return "false";
    else{
    $found=array_search($word,$exception);
        if($found)return "false";
        else{           $found=array_search($word,$exception4);
            if($found)return $word;
            else{
                //search in prefix          for($i=0;$i<sizeof($prefix);$i++){
                    $len=strlen($prefix[$i]);
        $pre=substr($word,0,$len);                  if($pre==$prefix[$i]){
$word=substr($word,$len);           break;
            }       }
```

4.0 Design and Implementation

4.1 Algorithm Execution

The best way to measure the work is to apply the algorithm dynamically within active application in order to get the results. When the structure is applied according to the programming language, it is very helpful to test and perform the work.

The implementation is as follows:

1. Install the application designed by PHP using VertrigoServ software to use it as web mining.

2. Use GUI of the searching in E-Library to get a physical result.

3.Process the result in communication with documents and databases.

4. Use applicable tools that support the work to be more powerful (corpus of Arabic texts).

5.Interface with the web mining by using special codes or local server to shift the program to the web.

The implementation of the E-Library application is as follows:

1. Run the application.

2.Choice the search section.

3.Enter the query in the text box.

4.Click the check box to get the root.

5.List the results of books.

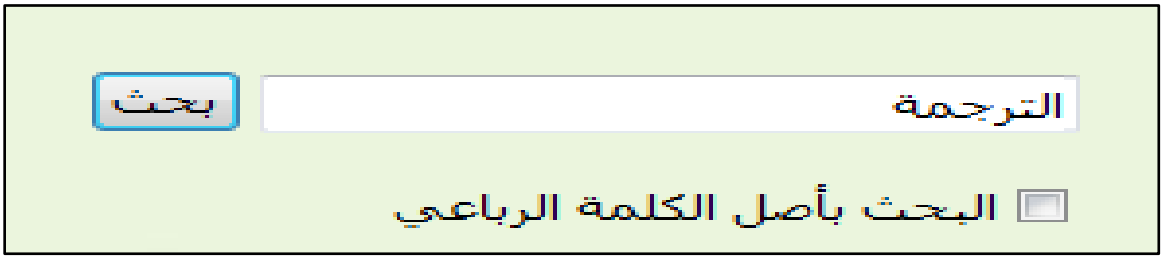

Figure 4.1.1: Tool Adopting for Quad Word

The figure below shows the results of the search in the books when use the ordinary search in the E-Library and before runs the quad algorithm that we adopt in the search of books by using tool in the application:

Figure 4.1.2: Results of the Ordinary Search

As we show in the top figure, the results of the number of search is shown in the top by books, titles, and authors. It found results regarding on the key word that written in the text box without any analyzes for word. By the way, any books library includes the search tool in order to help the user to get the relevant result of the query and help to expect the best ranking of result using specific steps to perform the results of search engine in the E-Library system.

But when we adopt the quadrilateral algorithm in the E-Library application it gave different results and ranking of the result that given in the ordinary search. The figure below describes the results when operate the algorithm for the same keyword and process the root of word:

Figure 4.1.3: Result of the Quad search

The E-Library application used to find the number of results and its ranking of the Arabic words based on user query. The tool adopted in the E-Library application used to get more results of the Arabic books related with the root of the quad word. The output helps to increase the information and reach our goals for the knowledge related with the user needs.

The quadrilateral tool provides rare services for the Quranic Arabic Language to support the user to analyze and check the root of the Arabic word regarding the structure of each word in the text box.

This work performs and processes the Arabic text after the word is written in the text box and the search button is clicked to get the result of the user query. Also, the check button tool will help to correct mistakes of quad roots in the Arabic text. It assists for the Arabic morphology and shows the root by given the frequency of results regarding on the user query. It's very helpful and allows the Arabic user to be more comfortable and closer with the result.

4.2 Steps to Run Program

The steps to operate the program are:

1. Write the Arabic word or copy past it from your collection into the text box.

2. Click the check button to run the quad algorithm.

3. Click the search button to get the results regarding on quad search.

4. Show the result of the books listed after search process of the information retrieval in the system.

5.0 Results and Discussion

5.1 Reliability of Survey

The stability of the results regarding the hypotheses is looking in the different between the ordinary search and quad search. The reliability statistics use Alpha Cronbach to measure the stability of the survey depends on the level of the test and the percentage of the high performance in the survey which ranges between 0 and 1.

The table 5.2 shows the cronbach alpha test result (0.787) with 11 numbers of questions view the value of reliability statistics.

Table 5.1.1: Reliability Statistics

Cronbach's Alpha	N of Items
.787	11

5.2 Confidence Interval Test

According to the analysis done between ordinary search and quad search, it can observe the Confidence Interval of the Difference (95%) in the ordinal search located between 1.34 and 1.66. In contrast, it located between 1.45 and 1.95 in quad search. So, the result of using algorithm for quadrilateral of the Arabic word is more effective than ordinary search regarding the results in analysis tests.

6.0 Conclusion

E-library is one of the interesting ideas followedfor study in the education. System includes a huge amount of information which is stored in special database or extracting from a corpus of documents. The E-library is a part of an information retrieval system. It provides methods to get information and increase knowledge.The quadrilateral algorithm has been adopted and integrated with New Approach for Extracting Quadrilateral Arabic Root and Pattern based Stemmer for Finding Arabic Roots.

References

[1] Al-Salman, A., Al-Ohali, Y., &AlRabiah, M. (2006). An arabic semantic parser and meaning analyzer. Egyptian Computer Science Journal, 28(3), 8-29.

[2] Alshalabi, R. (2005). Pattern-based stemmer for finding arabic roots. Asian Network for Scientific Information, 38-43.

[3] Boudlal, A., Belahbib, R., Lakhouaja, A., & others, (2011). A markovian approach for Arabic root extraction. The International Arab Journal of Information Technology, 8, 91-98.

[4] Gliem, J., &Gliem, R. (2003, October). *Calculating, interpreting, and reporting cronbach's alpha reliability coefficient for likert-type scales*. The Ohio State University Midwest research to practice conference in adult, continuing, and community education, Columbus.

[5] Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques. (Third ed.). USA: Elsevier Inc.

[6] Kanaan, G., AL-shalabi, R., & AL-Kabi, M. (2005). New approach for extracting quadriliteralarabic roots. ABHATH AL-YARMOUK, 14, 51-66.

[7] Kohonen, T., Kaski, S., Lagus, K., &Others, (2000). Self-organization of a massive document collection. IEEE TRANSACTIONS ON NEURAL NETWORKS, 11(3), 574-585.

[8] Landau, S., &Everitt, B. S. (2004). A handbook of statistical analyses using spss. Boca Raton London New York Washington, D.C.: Chapman & Hall/CRC Press LLC. DOI: www.crcpress.com

[9] Larkey, L., Ballesteros, L., & Connell, M. (2007). Light stemming for arabic information retrieval. Arabic Computational and Morphology, Springer, 221-243.

[10] Manning, C., Raghavan, P., &Schütze, H. (2009). An introduction to information retrieval. Cambridge, England: Cambridge University Press. Retrieved from http://www.informationretrieval.org/

[11] MCMILLAN, M. (2005). Data structures and algorithms using visual basic.net. New York, USA: Cambridge University Press. DOI: www.cambridge.org/9780521547659

[12] Norušis , M. (2007). Spss statistics base 17.0 user's guide. Chicago, USA: SPSS Inc.

[13] Rogerson, B. (2008). An evaluation of existing light stemming algorithms for arabic keyword searches. (Master's thesis, University of North Carolina).

[14] Stendahl, S., Andersson, A., &Strömberg, G. (2011). Web mining. (Department of Science and Technology, Linköping University)

[15] Thuraisingham , B. (2003). Web data mining and applications in business intelligence and counter-terrorism. The University of Texas at Dallas, Richardson, USA : CRC Press.

[16] Yao, J. T., & Yao, Y. Y. (2003). Web-based information retrieval support systems: building research tools for scientists in the new information age. Manuscript submitted for publication, Department of Computer Science, University of Regina, Regina, Saskatchewan, CANADA.

Mr. Said Mohammed Ali Al-Rashdi received Bachelor's degree in Computer in 2006. He is currently pursuing the Master Degree in Information Systems Major at University of Nizwa, Oman. His research interested in an Adopting Quadrilateral Arabic Roots in Search Engine of E-library System.

Dr.S.Arockiasamy, Head Department of Information systems, University of Nizwa. He received M.Sc, M.Phil and Ph.d in Computer Science. He is specialized in Applications of Image processing. He has published considerable number of research papers and articles in various leading International journals and International Conferences. He has been chief editor and two IT related magazines. He was also, leading common writer in a daily newspaper in India. Recently, he was awarded best professor in Information Technology by Asian Education Council.