

## **SEARCHING FOR ARABIC-BASED E-LEARNING WEB SERVICES: AN APPROACH TOWARDS USING SYNONYMS & DERIVATIVES**

*MAQBOOL AL-MAIMANI*

*Information & Computer*

*Department - Oman Air*

*Muscat, Sultanate of Oman*

Email: [maqbool@oas.com.om](mailto:maqbool@oas.com.om)

*AHMED AL-NAAMANY*

*College of Engineering,*

*Sultan Qaboos University,*

*Muscat, Sultanate of Oman*

E-mail: [naamany@squ.edu.om](mailto:naamany@squ.edu.om)

*OTTO SPANIOL*

*Informatik Department,*

*The University of Technology of Aachen,*

*Aachen, German*

E-mail: [spaniol@informatik.rwth-aachen.de](mailto:spaniol@informatik.rwth-aachen.de)

### **Abstract**

*Turning e-learning systems into Arabic-based web services represent number of challenge among which is how to effectively find the required e-learning services that are similar in meaning to the requested ones. This paper presents an overall description of a method which expands the search query by analyzing relationships among words that have similar meanings. The derivations of these words are obtained from their basic Arabic roots and the results are being analyzed to group those derivatives based on their absolute and relative meanings. This idea is being researched and the findings will be explained in future paper(s).*

### **Keywords**

Web Services, e-learning, Search Techniques, Synonyms, Derivatives of Arabic Words

## **I. INTRODUCTION AND BACKGROUND**

The Web Services technology has positioned itself as an ideal model to interoperate information systems regardless of their underlying technologies and platforms [1, 2, 3, 4]. This has enhanced the way we do business today since Web Services technology enabled the interaction of customers, suppliers and other business partners smoothly [2, 3, 5, 6, 7]. Like other electronic systems, e-learning systems have also realized the importance of using the internet and its services to enhance learning services to the learners [8, 9]. This include adopting web services as the enabling technology of the future [4, 10, 11, 12, 13].

Moreover, we also notice that there is a growing amount of multi-lingual electronic information on the web [14]. The global outreach of the web content has raised the importance of finding effective ways and methods in allocating the required information from the World Wide Web [14].

Searching for the right web service remains a challenge especially for Arabic-based contents [14, 15, 16, 17].

This paper presents an overall description of a method which expands the search query by analyzing different relationships among words that have similar meanings. The derivations of these words are obtained from their basic Arabic roots and the results are being analyzed to group those derivatives based on their absolute and relative meaning. This idea is being researched and the findings will be explained in future paper(s).

The paper consists of three main sections. Section II presents Web Services technologies and some essentials concepts. In section III e-learning and learning web services are introduced. Arabic language features and few search techniques are explained in section IV. The overall idea of our search approach is presented in sections V.

## **II. WEB SERVICES TECHNOLOGY**

Web services are set of technologies, which are based on Service-Oriented Architecture (SOA), that enable applications (including legacy ones) running on heterogeneous platforms at different geographical areas to interoperate with each other cheaply and easily [1, 3, 5, 6, 7]. For example, a mainframe running COBOL-based application can easily interact with a Java-based program without writing any interfaces between the two [3, 6, 7]. This is achieved by publishing program routines as services that can be invoked by the application requiring such a service [3, 5, 6]. Web Services are enabled by a set of standards and protocols that are based on XML technology. A service, which is a piece of software, is defined by a language called Web Services Description language (WSDL). The WSDL file contains details like: the service name, description and the address of service provider that can provide the service to the service consumers (requestors). The

service is published on the web in a registry using a protocol called Universal Description, Discovery and Integration (UDDI). UDDI is a technical specification that is used to describe, publish, discover (find) and integrate web services [1, 2, 3, 5, 6]. A service consumer queries UDDI to search for a required service and once found, UDDI will send the WSDL file to the service requestor. Based on the details given in the WSDL, the service requestor will contact the service provider and get the required service. All messages among service provider, UDDI and service consumer are exchanged using SOAP protocol. This interaction process is depicted in the Figure-1. Major companies like Sun, IBM, Microsoft, Oracle and SAP are adopting Web Services and accordingly building and releasing tools and applications that are Web Services enabled [3, 7].

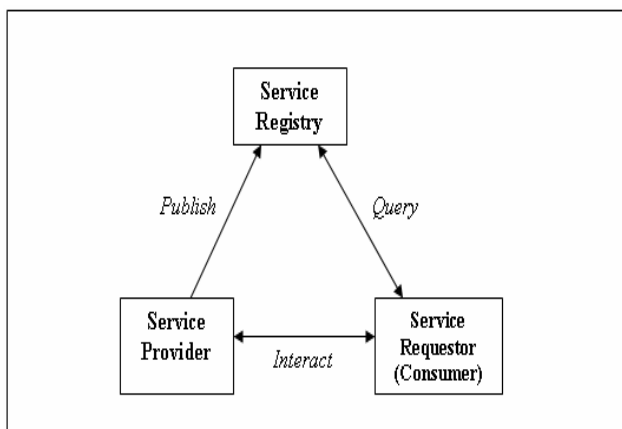


Figure-1: The Web Services Interaction Model

### III. WEB SERVICES & E-LEARNING

The internet played a major role in revolutionizing the Learning systems. One can easily notice that there are numbers of learning programs built for the Web along with dozens of learning portals. This has resulted in what is called today “electronic learning – e-learning” which enables learners to access learning materials and objects using the Internet [4, 8, 9]. Figure-2 depicts the components of an e-learning system. Learning Management System (LMS) is the core of an e-learning system. LMS is responsible for number of tasks like storing and managing learning contents in its database, scheduling learning events and managing learners’ profiles. The Run-time system makes the learning content available to learners. The creation and composition of learning contents are carried out by the Authoring part. [11, 12]. Learners, authors (teachers, trainers or instructional designers) are the main players of an e-learning system [8, 10, 11, 12, 18].

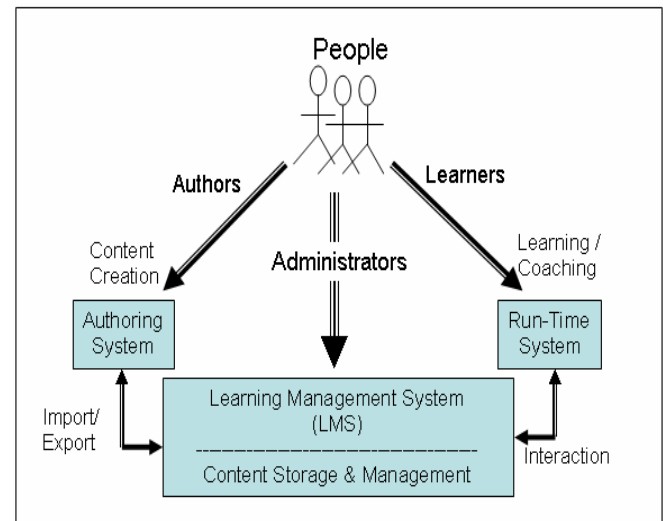


Figure-2: E-Learning System Components (Adapted from [11])

Both web services and e-learning share number of common features. This has easily enabled applying web services technologies onto e-learning domain. In e-learning we have providers who create course contents in form of Learning objects (LOs) and we have requestors who consume learning materials. E-learning functionalities and components can be realized as learning web services. LOs, accounting service for financial aspects, content authoring, authorizing service, learner registration, assessment service, monitoring service to monitor learners’ progress, chat service to allow learners to talk to their teachers, authentication service, content configuration to allow changes in learning contents, tutoring and exam are all can be viewed as web services [11, 12, 13]. These services can be described using WSDL and published as services in UDDI-like registries [12]. E-learning players can call these services and integrate them to create course contents to meet their needs. All course prerequisites will be monitored using the passing web services which checks if the learner cleared all required courses before taking the next one. If the learner fails a course, the passing web services will decide if the learner needs to repeat the course or take a similar course which consists of different LOs. All assistant, a learner may require while taking the course, will be provided by the help web service. It is important to mention here that learners who study the same course may not get the same LOs to work on but they will get similar ones. This is due to the fact that LOs are downloaded from different servers and different providers [12, 13].

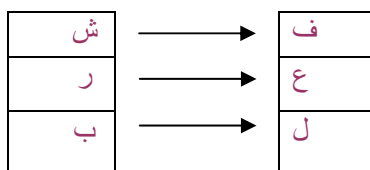
In view of the above, we can see that using web services in e-learning provides many benefits for learners, teachers, contents provider and authors. Among these benefits, LOs can be easily reused and combined with other LOs to create tailored courses. Moreover, since LOs can be logically or physically distributed on different servers located at different locations, this gives vendors the chance to specialize in offering services of their domain and interest. For instance, a vendor may specialize in financial related services. In addition, such distribution of services allows scalability,

maintainability and reusability of services. Also by defining LOs as services, such services can be easily updated and changed due to change in requirements or due to technology advancements. This helps the learner always to get materials that are up-to-date and according to the latest technology [12, 13].

#### IV. ARABIC LANGUAGE & FEW SEARCH TECHNIQUES

Arabic is one of the oldest languages in the world. It is the official and religious language of millions of Arabs and Muslims respectively around the world. Arabic is a very rich language that has number of features which represent in some cases challenges for Information Retrieval (IR). Some of these features and challenges are [14, 15, 16, 17, 19, 20, 21, 22, 23]:

- Arabic consists of 28 characters that can be written also in different shapes based on vowels, marks and connectivity with each other. [21].
- Arabic words are written from right to left. Arabic characters have diacritical marks called *Fathah*, *Dammah* and *Kasrah* which decide how a character should be pronounced. Double letter is represented by using what is called *Shaddah* [17]. The *Shaddah* is shown like 'w' on top of a word. For example, the word فَعَّال (which means *effective*) has *Shaddah* on the 2<sup>nd</sup> character which indicates that this character is a double character.
- There is a special character called Hamza which is written during official writing and omitted in normal occasions.
- The shape of Arabic character depends on the position of the character within the word and the one before it and after it. This makes printing, displaying and searching more challenging than English.
- Arabic words consists of three main categories : Nouns (both Adjectives & Adverbs), Verbs and particles.
- The basic form of an Arabic word (after removing all its prefixes and postfixes) is called the root of the word. The root can be of three letters (trilateral), four letters (quadrilateral) or five letters (peritalliteral). Any further analysis / changes on the root will result in loosing the meaning of the word. The Arabic root of any word is balanced (mapped) to a standard form of three, four or five characters which are (fa'ala (فعل), (fa'lala (فعلل), (fa'lalal (فعللل) respectively. For example, the word شرب (which means drink) is balanced as فعل, as can be seen below that each character of the root word شرب has a corresponding character in the balancing form فعل. [14, 16, 21, 22, 24].



- Nouns are prefixed and post fixed by few letters. Most nouns are prefixed with 'ال' ('al').
- There are three types of plurals in Arabic: masculine, feminine and broken plurals [25]. Plurals are formed by adding suffixes to the noun or changing its pattern [25]. In case of broken Plural, the singular noun is completely altered to arrive to its plural form. In this case, one cannot be certain of how suffixes and prefixes will be in such plurals. Broken plurals don't resemble singular form as compared to irregular plurals in English language [19, 25].
- There is no space between a word and its pronoun. For example, 'her book' is written as كتابها where كتاب means 'book' and ها means 'her.' 'We write' in Arabic is written as نكتب where كتب means 'write' and ن stands for 'we', as an another example [25].

Most of the above features represent challenges while searching for Arabic-based content. Studies for retrieving Arabic information has been started since 1980s [17]. Few search techniques (which are also applied for Arabic searches) are listed below [14, 16, 17, 21, 26].

1. **Literal Search:** This is the basic search mechanism which matches the terms as entered by the user. This method has a very low success rate since the searched documents may not have terms written the way they are entered.
2. **Stemming:** This process applies some normalization on the words that need to be searched. Such normalization includes removal of all prefixes, postfixes, unification of *Hamza* and standardization of other signs. By this process all words will be reduced to a common form [15, 17, 19, 21, 22, 26, 27].
3. **N-Gram:** In this method, a length of n consecutive characters is extracted from a word. The commonly used values for n are 2 or 3. The idea here is that there is a high probability that similar words will have a high proportion of n-grams in common [17, 19, 21, 26, 28, 29].
4. **Soundex:** This search algorithm returns words that sound similar to the entered word. This method was initially designed by The United States Census Bureau in 1935 [17]. Soundex algorithm keeps the first characters of the word and replaces succeeding characters with numeric values. Character w, y and h and vowels are ignored. Constants that are adjacent are represented once. Based on this, the Soundex algorithm generates a code of length 4 (starting with character and followed with a number). This code is compared with Soundex code of other words and those similar words that have similar Soundex codes are returned to the user. There are different versions of Soundex and there are few efforts to develop Arabic Soundex algorithm [17, 28, 29, 30].

## V. OUR APPROACH

Alta Vista, Google and Yahoo are some example of well-known search engines that we use on daily basis for searching the web. Though these engines can handle other languages, they are mainly designed for effectively searching English-based contents [14, 15, 16, 17, 21]. There are number of efforts to develop Arabic search engines using various techniques like the ones described above. However, there is a lot that needs to be done in this area.

Our search approach, whose general idea is described here, is based on two major features of Arabic: word synonyms and derivatives. The Arabic language is rich of synonyms and usually one word has many synonyms that have either absolute or relative meaning. An absolute synonym is where two words have exactly the same meaning; whereas, a relative synonym is where two words have close meaning [16]. Table-1 gives examples of few groups that have synonym words. The second aspect of the research is based on the fact that each root of a word can have many derivatives, as per Arabic linguistic rules [24]. Table-2 lists major derivative forms that an Arabic word can have. Our approach will start with finding derivations of few synonym groups, as per the derivative forms shown in table-2 [24]. The results will be analyzed in depth to find out how the words are related in each group. Within one group we may find that there are derivatives that are absolute synonyms; whereas, we may find within the same group that there are derivatives that are relatively close in meaning. In addition, there might exist derivatives that are far in meaning from others. An in-depth analysis will be carried on these lines with multiple groups of synonyms and results will be compared and analyzed to arrive to some general patterns among these groups. Sets of words, within a group, that are very close in meaning will be assigned a higher percentage than another set, within the same group, containing words that are less close in meaning. Based on conclusions that will be drawn at the end of the research, Fuzzy Logic and Neural Networks techniques will be applied to encode such patterns in an intelligent system. Future searches will work as follows: a user will enter word(s) (or a web service name) for a search. The word will be converted back to its root and all synonyms to the root will be found. All derivatives of these synonyms will be extracted and based on the Neuro-Fuzzy logic, the words that have similar meaning will be generated to expand the query of the user. The new expanded query will be used to find all related web services using the UDDI registry. Table-3 illustrates an example of two groups of synonyms with their derivatives.

**Table-1: Groups of Synonym Words**

Group-1 (means to Stand)- Absol Synonyms نهض - قام - وقف	Group-2 (means to Collect) - Relative Synonyms جمع -- لَفَّ - ضَمَّ
Group-3 (means to Obtain) - Absolute Synonyms حاز - حصل -- نال	Group-4 (means to Walk) - Absolute Synonyms سار - مشى - خطى

**Table-2 : List of Derivative Forms**

<u>Name Derivative Forms</u>
اسم الفاعل / اسم المفعول / صيغة المبالغة / الصفة المشبهة / اسم التفضيل / اسم التعجب / اسم المكان اسم الزمان / اسم الآله / اسم الحال / الاسم المبني للمجهول / شبه المجموع / الاسم الصحيح / الاسم المقصود / الاسم المحدود / الاسم المنقوص / جمع التكسير / جمع المذكر والمؤنث السالم / التصغير / النسبة / جموع الكثرة / جموع القلة / منتهى الجموع / الاسم المنتهي بياء مشدودة / الاسم المضاعف / الاسم المركب / الاسم الأجوف / الاسم النيدل / الأبدال
<u>Verb Derivative Forms</u>
تصريف الأفعال (المضارع والماضي والأمر) وفقا للضمان أنا - هو - هي - هما - هم - هن - أنت (المذكر) - أنت (المؤنث) - أنتم - أنتن ( / توكيد الفعل / المصدر الصناعي / المصدر الميمي / مصدر المرة / مصدر الهيئة / الفعل المضاعف / الفعل المعتل / الفعل الناقص / الفعل الأجوف / الفعل المهموز

**Table-3: Two groups of Synonyms & Derivatives**

Group-1 (derivatives of words synonyms for the word -'go')			Form	Group-2 (derivatives of words synonyms for the word -'run')		
راح	ذهب	سار	فعل	أسرع	ركض	عدا
يروح	يذهب	يسير	يفعل	يسرع	يركض	يعدو
تروح	تذهب	تسير	تفعل	تسرع	تركض	تعدو
روحة	ذهبة	سيرة	فعلة	سرعة	ركضة	عدوة
تراوح	تذاهب	تساير	تفاعل	تسارع	تراكض	تعدى
مروح	مذهب	مسير	مفعل	مسرع	مركض	معدى
رواح	ذهاب	سيار	فغال	سراع	ركاض	عداء
روح	ذهب	سير	فعل	سرع	ركض	عدى
مرواح	مذهاب	مسيار	مفعل	مسراع	مركاض	معداء

## VI. CONCLUSION

Till date Web services are viewed as the best technology for interoperating various systems in all domains including e-learning. Developing proper search techniques to find the required web service is very essential. This paper presented an overall description of a search method that is being studied and researched in-depth. This research may result in opening another window for Arabic-based searches. Our findings will be presented in future paper(s) and all open issues and challenges (covering Arabic linguistic structure) will be published for futures researches.

## REFERENCES

- [1] Zhengfang Xu, Zheng Yin and Abdulmotalen El Saddik, "A Web Services Oriented Framework for Dynamic E-Learning Systems", 2003, Published by the IEEE Computer Society
- [2] Minder Chen, Andrew N K Chen, Benjamin B M Shao, "The Implications and Impacts of Web Services to Electronic Commerce Research and Practices"
- [3] Doug Kaya, "Loosely Coupled: The Missing Pieces of Web Services", (USA, RDS Pres, 2003).
- [4] Xiaofei Liu, Abdulmotalen El Saddik and Nicolas D. Georganas, "An Implementable Architecture of An E-Learning System", 2005, Published by the IEEE Computer Society
- [5] Douglas K. Barry, "Web Services and Service-Oriented Architecture: The Savvy Manager's Guide", (San Francisco USA, Morgan Kaufmann Publishers, 2003)
- [6] Ethan Cerami, "Web Services Essentials", (Sebastopol, California, USA, O' Reilly & Associates Inc. 2002)
- [7] H. Peter Alesso & Craig F. Smith, "Developing Semantic Web Services", (Massachusetts, USA, A K Peters Ltd, 2005)
- [8] Marc J. Rosenberg, "e-learning", (USA, McGraw-Hill 2001)
- [9] Jay Cross, "An Informal History of eLearning", On the Horizon, Volume 12, Number 3, 2004, pp 103-110.
- [10] Ljiljana Stojanovic, Steffen Staab, Rudi Studer, "eLearning based on the Semantic Web"
- [11] Victor Pankratius, Olivier Sandel and Wolfried Stucky, "Retrieving Content With Agents in Web Service E-Learning Systems."
- [12] Gottfried Vossen and Peter Westerkamp, "E-Learning as Web Service (Extended Abstract)", 2003, ", Published by the IEEE Computer Society.
- [13] Gottfried Vossen and Peter Westerkamp, "Turning Learning Objects Into Web Services", April, 2005
- [14] Haider Moukdad and Andrew Large, "Information Retrieval from Full-Text Arabic Databases: Can Search Engines Designed for English Do the Job?", 2001
- [15] Haider Moukdad, "Lost In Cyberspace: How Do Search Engines Handle Arabic Queries?"
- [16] Hyder K. Al Ameen, Shaikha O. Al Ketbi, Amna A. Al Kaabi, Khadija S. Al Shebli, Naila F. Al Shamsi, Noura H. Al Nuaimi, Shaikha S. Al Muhairi, "Arabic Search Engines Improvement: A New Approach using Search Key Expansion Derived from Arabic Synonyms Structure", Published by the IEEE Computer Society, 2006.
- [17] Syed Uzair Aqeel, Steve Beltzel, Eric Jensen, David Grossman and Ophir Frieder, "On the Development of Name Search Techniques for Arabic", Illinois Institute of Technology, Published in Journal of American Society for Information Science and Technology, Wiley Periodicals, 2006.
- [18] Victor Pankratius and Gottfried Vossen, "Towards E-learning Grids: Using Grid Computing in Electronic Learning", Published by IEEE, Oct 2003
- [19] Jinix Xu, Alexander Fraser and Ralph Weischedel, "Empirical Studies in Strategies for Arabic Retrieval", August 2002
- [20] William Black, Sabri Elkater and Piek Vossen, "Introducing the Arabic WordNet Project."
- [21] Mohammed Aljlal and Ophir Frieder, " On Arabic Search: Improving the Retrieval Effectiveness via a Light Stemming Approach", November 2002
- [22] Abdur Chowdhury, Mohammed Aljlal, Eric Jensen, Steve Beitzel, David Grossman and Phir Frieder, "Linear Combination Based on Document Structure and Varied Stemming for Arabic Retrieval", Illinois Institute of Technology, 2002
- [23] Taleb Ahmed and Julius Mungo, "On the Difficulty to Design Arabic E-Learning System in Statistics", January 2006.
- [24] Dr. Mahmood Matraji, "في الصرف وتطبيقاته" (Arabic Derivation & its Applications), (Beirut, Lebanon, دار النهضة العربية للطباعة والنشر, 2000)
- [25] M. Aljlal, O. Frieder and D. Grossman, " On Bidirectional English-Arabic Search."
- [26] F. Cuna Ekmekcioglu, Michael F. Lynch and Peter Willett, "Stemming and N-gram Matching for Term Conflation in Turkish Texts", Oct 1996.
- [27] T. Rachidi, O. Iraqi, M. Bouzoubaa, A. Ben El Khattabi, M. El Kourdi, A. Zahi and A. Bensaïd, "Barq: Distributed Multilingual Internet Search Engine with Focus on Arabic Language", published by IEEE, 2003
- [28] David Holmes, Samsun Kashfi and Syed Uzair Aqeel "Transliterated Arabic Name Search."
- [29] ByungRae Cha, " Host Anomaly Detection Performance Analysis Based on System Call of Neuro-Fuzzy Using Soundex Algorithm and N-gram Techniques", Published by IEEE, 2005
- [30] Moawia Elfaki Yahia, Randa Eltayeb Abd Elhafez, Tumadir Badawi Elsherif and Omima Nasser Osman, "Intelligent Soundex Function for Arabic Names", Published by IEEE 2004.