



A Novel Method to Evaluate Romanization Systems:

The Case of Romanizing Arabic Proper Nouns

Mohammed N. AI-Kabi Computer Science Department Faculty of Information Technology Zarqa University P. O. Box 2000 13110 Zarqa -Jordan malkabi@zu.edu.jo Hanan N. Abu Obied CIS Department IT & CS Faculty Yarmouk University Irbid, Jordan hanannd@yu.edu.jo Izzat M. Alsmadi Computer Science Department University of New Haven West Haven, CT 06516, USA ialsmadi@newhaven.edu Maryam S. Nuser CIS Department IT & CS Faculty Yarmouk University Irbid, Jordan mnuser@yu.edu.jo

Abstract: The transliteration of Arabic proper nouns to other languages is usually based on the phonetic translation of these nouns into their phonetic Latin counterparts. Most of the dictionaries do not include most of these nouns, although some may have meanings. Transliteration is essential generally to Natural Language Processing (NLP) field and specifically to machine translation systems, cross-language information-retrieval systems and Web search engines, since most of the submitted queries are proper nouns. Romanization also known as Latinization that refers to the representation of names and technical terms with Roman (Latin) alphabets. Romanization is accomplished based on a number of methods. Machine Romanizers are either based on a dictionary or on rules to Romanize different proper nouns. The Romanization process is not trivial due to many problems, so normally there are different Roman variations of a single Arabic proper noun that cause many problems to search engines and databases. This study is based on a dataset consisting of around 5000 Arabic proper nouns as a case study, where this dataset accumulated through years as a result of indexing the authors of Arabic books at the library of one of the public universities in the Middle East. Many methods are presented to evaluate Romanization systems. In this paper, we presented a new automated evaluation method called, "Back Romanization" to automatically evaluate the effectiveness of different Romanization languages used.

Keywords: Romanization System Evaluation; Arabic Romanizers Evaluation; Rule-based Romanization; Dictionarybased Romanization; Arabic names Romanization.

1. Introduction

The Arabic language is the official language in twenty seven countries, and one of the six official languages used in the United Nations (UN) [1][2]. It is the most popular Semitic language and a native language of more than 422 million people and used by more than 1.5 billion Muslims [2][3]. It is written in a cursive way from right to left. The spread of Islam leads to the increase of the usage of Arabic language these days by around 1.5 billion Muslims worldwide. The 28 Arabic alphabets are adapted for the writing of different languages like Urdu, Pashto, Ottoman, Persian, Swahili (in East Africa), Uyghur (in China and Central Asia), Kurdish (in Iraq and Iran), Sindhi ...etc. Many of the Arabic letters look similar and distinguished from one another by points (dots) which exist either above like the Arabic letter (xaa', " \dot{z} ") or the point could be below as in the Arabic letter (jiim, "כ"), or the Arabic alphabet could be free from any dot as in the letter (haa', " τ "). Arabic letters are not capitalized, so it is not easy to distinguish proper nouns in Arabic. Arabic alphabets have different shapes based on their position (Beginning, middle, separate and an end) within different Arabic words. Also, there are Arabic alphabets with different shapes regardless of their

position like the first Arabic letter (Alif, "ا، أ، إ، أ، ى") with five different shapes. In addition to these letters (consonants), the Arabic language uses diacritical marks like Fatha mainly in holy books [4][5], are written only above or below the body of the Arabic word. Eight Arabic diacritical marks are used beside the 28 Arabic alphabets mentioned before [6]-and they are: Fatha ´, Damma ['], Kasra ['], Fathatayn ['], Dammatayn ['], Kasratayn ['], Shadda ['], and Sukun[°] [7]. (hamza, "•") is an additional character that has no corresponding Roman letter when it is used separately, but it can be combined with other Arabic letters (long vowels) like (Alif, "أ، إ"), (Yaa', "ى") and (waaw, "ؤ") ("أ، إ") [6]. Proper nouns are usually transliterated and not translated depending on their meanings. Therefore, transliteration process means converting text written in one writing script to a text written by another writing script depending on the phonetics of the writing scripts of the source and target texts. In addition, the transliteration process is beneficial to those who cannot read the Arabic script [6].

The translation from Arabic to English, Spanish, Portuguese, French, Swedish, Dutch, Bahasa Indonesian, Turkish, Tagalog, Malaysian (Rumi), etc., usually faces the problem of how to Romanize (transliterate) Arabic proper nouns to their counterparts in any one of the languages that adopt Roman alphabets (Latin alphabets). Kindly note that Romanizing an Arabic proper noun to English, for example, may be different from Romanizing the same proper noun to French, i.e., Romaniziation of the Arabic female name (Shereen, "شيرين") to Shirin in English and to Chirine in French, also the Romaniziation of the Arabic name of the capital of Lebanon (Beirut, "بيروت") to Beirut in English and to Beyrouth in French. The transliteration process is essential in many fields of our modern life, within banks, airlines, academic establishments, residency offices, home offices, media, libraries, machine translation...etc. This process of converting Arabic proper nouns to their phonetics equivalent is called Romanization, since this process converts Arabic alphabets to Roman alphabets [8]. On the other hand, Arabization refers to the opposite process of the Romaniziation, where a proper noun written in Latin script is converted to its phonetic equivalent in Arabic.

Two main methods are used to automate the Romanization process: Dictionary-based and Rulebased. Dictionary-based Romanization method; also called database method, is based on storing the source Arabic proper nouns in a database with the Romanized proper nouns as suggested by experts in this field. This leads to building a highly effective system for all Arabic proper nouns stored in the database, but with a limited capability since it cannot Romanize Arabic proper nouns that are not already stored in the database. Dictionary-based Romanization systems are simple since they are really lookup tables from a programming point of view, which are based mainly on human effort to Romanize Arabic proper nouns. Therefore, the system that has a larger number of Arabic proper nouns with their Romanized proper nouns is considered better than others with fewer Arabic proper nouns. The Rulebased approach is the second main method which is also known as spelling transliteration method. Rulebased approach converts Arabic proper nouns to Romanized proper nouns by replacing each Arabic alphabet in the source Arabic proper noun with corresponding equivalent phonetically Roman alphabet(s).

The difficulties that face the Romanization of Arabic proper nouns are summarized by the lack of standards before 2004, and even when these standards exist there is no guarantee that employees in charge will follow these standards to Romanize Arabic proper names [9]. A study conducted by Jack [10] refers to difficulties of Romanizing Arabic proper nouns. These difficulties as presented in Jack [10] study are due to: Arabic/English differences of phones and graphemes, ambiguity of the Arabic language, where Arabic proper nouns are not easily identified as proper nouns like in English (Capitalization), and a good portion of Arabic proper nouns have meanings and in this case the system in use has to determine whether to translate or Romanize. Therefore, a number of studies are conducted to show how to extract Arabic proper nouns, and one of these is conducted by Al-Shalabi et al. [11]. Arabic and English languages are based on a different number of consonants, therefore the mapping process will not be straight forward. Mapping process of Arabic consonants to their Latin counterparts may vary according to the context, i.e., Arabic consonant (siin, (ω) may be mapped either to the Latin letter C or S. Arabic consonant (baa', \rightarrow) may be mapped either to the Latin letter B or P, Arabic consonant (kaaf, ك) can be mapped either to the Latin letter K or C, and in this case we have two Arabic letters (siin, س) and (kaaf, ك that could be mapped to the Latin letter C [12]. In some cases, the mapping of an Arabic letter needs two Latin counterparts like within the Arabic alphabet (shiin, ش) which is mapped to sh or ch. Also, there are two Arabic alphabets (thaa', ث) and (dhaal, ن) which mapped to two Latin counterparts (th). are Furthermore some of the Arabic proper nouns are transliterated before centuries into a phonetically different version from the Arabic source like (Jerusalem, الأردن, Jordan, الأردن), (Babylon, إبابل), (Tripoli, طرابلس), (Aleppo, حَلَب), (Damascus, دمشق), (Sidon, اصيدا, (Egypt, مصر), (Alexandria, الإسكندرية), (Cairo, المغرب), (Algeria, الجزائر), (Morocco, المغرب), (Somalia, الصومال), (Tunisia, تونس), (Emirates, الإمار ات), etc.

There are many evaluation methods for evaluating different Romanization systems. This study proposed a new method called Back Romanization as an automatic evaluation method. This is a bi-directional system with the ability to Romanize the Arabic proper nouns which consist of Arabic alphabets to Roman proper nouns. The system can also convert back the resulted Roman proper nouns to their equivalent Arabic proper nouns. The evaluation which is based on Back Romanization of automatic Romanizers considers a Romanization process successful if the output of this bi-directional process succeeds in outputting the original Arabic proper noun once again. The Back Romanization method can be used to evaluate different Romanization/Transliteration systems regardless of the language(s) they used. Therefore, one of the merits of this method it is a language independent method.

The rest of this paper is organized as follows: Section 2 presents a summary to a number of studies related to Arabic Romanization (transliteration). Section 3 presents the methodology adopted in this study. Section 4 presents the results of the tests conducted on the proposed method (Back Romanization). Section 5 exhibits the conclusions and future plans to improve this study.

2. Related Work

Researchers have exploited various methods to evaluate Romanization/Transliteration methods, so in this section an overview of previous studies related to Romanization, Romanization systems, and their evaluation is presented. Although the proposed method "Back Romanization" in this study is a language independent, but it is based on a dataset of Arabic proper nouns; therefore studies related to Arabic Romanization/Transliteration are presented in this section, beside those that include evaluations of different Romanization/Transliteration systems.

Thomas Edward Lawrence (Lawrence of Arabia) is a pioneer who used the Romanization of Arabic proper nouns extensively in 1926 when he sent 130,000 Arabic words manuscript about the Revolt in the Desert [13]. This manuscript was full with inconsistent Romanized Arabic proper nouns, i.e., (Jeddah, "جَدّة") city in western part of Saudi Arabia mentioned in Lawrence manuscript once as Jeddah as well as Jidda [14]. Furthermore [14] refers to percentages of different variations of the name of Prophet (Muhammad, "محمد") typed in by Google Search Engine users as follows: (Muhammad 41%), (Mohammed 32%), (Mohamed 25%), and (Mahomet 2%). Also [14] presents different percentages of use with Google Search Engine to the name of the late Libyan leader: (Gaddafi 72%), (Qadhafi 16%), (Gadafi 8%), (Gadafy, 2%), and (Qadhdhafi 1%).

Afterward, International Convention of Orientalist Scholars organized by German Oriental Society in Rome witnessed another attempt for Arabic Romanization in 1936 [13]. The Romanization System for Arabic was adopted by United States Board on Geographic Names (BGN) in 1946, and this was applied by the Permanent Committee on Geographical Names (PCGN) for British official use in 1956 and used in many Arab countries [15]. Another Romanization System for Arabic was adopted by French National Geographic Institute (IGN) in 1963 [16] [17] [18]. British Standards Institution (BSI) issued a standard for transliterating Arabic alphabets called BS 4280 in 1968, which needs to be purchased from BSI to be used [14]. Romanization of the Uighur, Urdu and Persian that use Arabic script and written from right to left was published in the eighth United Nations conference on the standardization of Geographical Names [19].

The First Arab Conference on Geographic Names (ACGN) was held in Beirut-Lebanon in 1971 and was attended by a number of Arab geographic and cartographic institutions beside experts from the United Nations. ACGN conference discussed the transliteration of Arabic geographic names into their Latin counterparts. One of the main achievements of this conference was a recommendation of a unified

Romanizer called Beirut System to be used by United Nations [13], and approved on the sidelines of the second United Nations Conference on the Standardization of Geographical Names (UNCSGN) in London (1972) [18][20][21]. An initial report version 1.3 (March 2000) was published by UNGEGN Working Group on Romanization Systems refers to approval of the 1971 unified Romanizer which is called Beirut System in 1972. Arab League was also urged to organize a specialized Romanization conference to discuss all aspects of the Romanization process and its problems with their possible solutions [22]. The I.G.N. system is released in 1973, and it is also called variant B of the Amended Beirut system 1971. This system is presented within the 17th session of UNGEGN in New York, 1994 [18]. Arab Division of Experts on Geographical Names (ADEGN) accepts the amendments to the "Modified Beirut System" during the Eighth United Nations Conference on the Standardization of Geographical Names (UNCSGN) held in Berlin in 2002 [18] [23].

Most of submitted queries to Web search engines and to Cross Language Information Retrieval (CLIR) systems are proper nouns. Therefore, Larkey, AbdulJaleel, and Connell [24] study explored the problem of retrieving proper nouns within CLIR, and focus on the importance of proper nouns within CLIR, where the effectiveness of CLIR system was degraded tremendously when it is based on bilingual lexicons that lack proper nouns. A number of different sources of English to Arabic proper nouns translation were tested. Larkey, AbdulJaleel, and Connell [24] study concludes that a combination of static translation resources and transliteration provides a successful solution to the problem of retrieving proper nouns within CLIR.

A new method for automatic learning transliteration system which is based on a sample of name pairs in Arabic and English is presented by AbdulJaleel, and Larkey [25]. The evaluation tests on their system reveal that the accuracy of the system depends mainly on the size of training data, so enlarging training data leads to an increase in the effectiveness of the system.

Hassan and Sorensen [26] present in their study an integrated approach for named entity (NE) translation which is based on a single framework composed of three combined modules: phrase-based translation, word-based translation, and transliteration modules. This integrated approach aims to benefit from the pros of each of three adopted modules while avoiding their pitfalls. Also a new approach for aligning NEs across parallel corpora is presented, which automatically extracts new NEs translation phrases.

Jack Halpern is one of the active researchers in the field of Arabic Transliteration, Romanization and Arabization. He conducted a series of studies in this field. Halpern [27] conducted a study that aims to accomplish two goals: Romanizing Arabic unvocalized proper nouns into various Romanizations systems and the Arabization of non-Arabic proper nouns specifically the CJK (Chinese, Japanese, and Korean) scripts, where two algorithms are presented to accomplish the two goals of his study. The author claims that this is a pioneering effort to automatically arabize the Chinese, Japanese, and Korean script. The author emphasis on the importance of linguistic knowledge to transliterate, for example, the Japanese script into Arabic. The accuracy metric is used to evaluate the algorithms, where Japanese-to-Arabic transcription yields a very high accuracy rate.

Halpern [28] study exhibits the challenges of orthographical complexities of Chinese, Japanese, Korean (CJK) and Arabic. The lack of standard orthography and irregularity leads to the complexity. Using linguistic knowledge supported by large-scale lexical databases will lead to overcoming these complexities according to the author. He shows that using a comprehensive, up-to-date lexical resources leads to achieving high accuracy in disambiguating and processing orthographic variants.

Halpern [29] study presents spelling variations of Arabic proper nouns, with an emphasis on Romanized Arabic proper nouns variations. To disambiguate these different variants of the Romanized Arabic proper nouns, the author proposes to use statistical modelling techniques such as Hidden Markov Model (HMM), and lexical databases of Arab names and their name variants in Arabic and Roman.

Fattah and Ren [30] exhibit in their study a new system with a number of approaches to extract proper nouns from two different English-Arabic parallel corpora using Dice similarity coefficient, Longest Common Subsequence Ratio (LCSR), Edit Distance and Longest Common Prefix Ratio (LCPR). The proposed system exploits the percentage of repetition factor of the English-Arabic proper noun pairs to enhance the overall performance of the system in terms of precision. Fattah and Ren's system is characterized with its ability to extract low-frequency transliteration pairs.

Alkharashi [31] study exhibits an infrastructure specified to help and simplify processes related to the generation, recognition, translation and transliteration and correction of Arabic person named entities. The infrastructure is based on two-dimensional person name map table that maps Arabic person name to its pattern which acts as a training set and helps to identify some affixation and gender characteristics. After identifying the pattern of the Arabic person name, it will be used as a model for person name to facilitate the process of person named entity recognition, correction, and transliteration. Lawson conducted a study to evaluate the six Arabic transliteration systems adopted by American Library Association and the Library of Congress and other four alternate Arabic transliteration systems. Different metrics are used in Lawson studies like phonetic and spelling accuracy, usability, and the adherence to not using non-native diacritics. These evaluation tests reveal that Qalam is the best Arabic transliteration systems, followed by ALA-LC and the other four systems. Letters and symbols used by Qalam system can be displayed by any Online Public Access Catalog (OPAC) [32] [33].

A comparison between dictionary-based approach and rule-based approach to transliterate Arabic names to English is conducted by Abu Obied, Nuser, and Al-Kabi [34]. Those authors use more than 5000 unique Arabic names in their study to identify the effectiveness of the above two approaches. They conclude that dictionary-based approach as expected is more effective than rule-based approach.

3. Methodology

This study is based on a dataset of Arabic author names collected from Library system of one of the public university Libraries. The size of the collected dataset used in this study is relatively small since it contains around 5,000 Arabic author names.

Transliteration systems have been evaluated using many methods, and investigated from different perspectives, and have been evaluated using many criteria and attributes (usability, accuracy, using diacritics or not, etc). Some of these methods evaluate the quality of the output from the perspective of people speaking both languages, where some methods evaluate the rules adopted by different systems to transliterate different proper names. On the other hand, some quantity methods count the number of names the system is able to transliterate correctly out of the total names. The quantity method used in this study is based on the following steps:

- 1. An Arabic proper name is automatically transliterated into Roman letters using a Romanization system.
- 2. The output of the first step (Romanized proper name) is back transliterated into Arabic letters using the same system.
- 3. The outputs of the second step are compared with the corresponding source Arabic proper names which were used in the first step. The system counts the number of fully matched names relative to the total number of Arabic proper names inputted to the system.

The above three steps are summarized in figure 1.



Figure 1. Flowchart of Back Romanization Method.

4. Experiments and Results

Dictionary-based approach and rule-based approach are the two main approaches used to Romanize Arabic proper nouns. The dictionary-based approach is characterized by its accuracy and its limitation to the size of proper nouns stored in. The rule-based approach is characterized by its capability to Romanize large number of Arabic proper nouns (Comprehensiveness) while it is less accurate than dictionary-based approach. It is known that human judgment is generally subjective, but it is more accurate than the objective automatic evaluation methods. Therefore, we rely on an expert in this field to evaluate our method. Our tests show that the Rule-based Romanization system used in this public university is capable of romanizing 85% of our 5000 names correctly as shown in Table 1. This system adopts the rules of the Library of Congress (LOC) to Romanize different Arabic proper nouns.

Table 1. Accuracy	Percentages of	Rule-based System
-------------------	----------------	-------------------

	Correctly	Incorrectly	
	Romanized	Romanized	
Arabic Proper Nouns	85%	15%	

Fattah and Ren study [30] presents the following three measures (Precision, Recall, and *F*-Measure) to evaluate the extraction of transliteration proper-noun pairs from parallel corpora based on different similarity measures between the English and Romanized Arabic proper nouns:

$$Precision = \frac{Correct}{Correct + Wrong},$$

$$Recall = \frac{Correct}{Correct + Missing},$$
$$F = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)}.$$

We found that these three measures cannot be used in this study, due to the fact that the precision is equal to accuracy, and the use of rule-based approach leads to makes the *Missing* variable used in the above *Recall* always equal to Zero (*Recall* =1). Accordingly the F measure cannot be computed.

Furthermore, the accuracy is used mainly to evaluate different Romanization systems, since in such cases we do not have relevant results, but we have either correct or incorrect Romanized outputs. Therefore measures like *Recall*, *Precision*, and F_1 cannot be used [26][35].

Furthermore, the conducted tests on the Back Romanization method that is presented in the previous section shows that the accuracy of this method is lower than the accuracy of the Rule-based Romanization system that is used at this public university, as presented in table 1. Table 2 shows the accuracy of the Back Romanization method as an automatic evaluation method.

Table 2. Ac	curacy of Bac	k Romanization	n Method.
-------------	---------------	----------------	-----------

	Correctly	Incorrectly
	Back	Back
	Romanized	Romanized
Arabic Proper Nouns	80%	20%

Our investigations lead us to the main reason behind errors committed by the Back Romanization system. We discover that the inability of the used system to restore back the Arabic letter ('ayn, " ξ ") in step 2 of the methodology, and this is due to the lack of corresponding Latin letter to the Arabic letter ('ayn, " ξ "). This incapability leads to a degrade in the accuracy of Back Romanization method. The following example demonstrates the cause of one of the errors in this method.

Example This example demonstrates a failure case for the above method used in this study to generate back the original Arabic name (Ebraheem Alwan, " ابراهيم ") as shown in table 3.

Table 3.	English Ron	nanization &	Back-Rom	anization

Original Arabic Name	English Romanization	Back - Romanization
ابراهيم علوان	Ebraheem Alwan	ابر اهيم ألوان

Table 3 shows clearly that the adopted method fails to generate back the Arabic alphabet ('ayn, " ξ ") and instead generate the Arabic alphabet (Alif, "'"). To those who are not Arabic native speakers these are Arabic alphabets seem phonetically similar alphabets, like the phonetics of English letters (B) and (P) to non native English speakers.

5. Conclusion and Future Work

Two approaches are used to transliterate Arabic proper nouns to English, French, German, etc. The first approach is called dictionary-based approach which is characterized by its high accuracy, but it is constrained by the size of the proper nouns stored in the system. The second method is called rule-based which is characterized by its capability to transliterate any Arabic proper noun, but the accuracy of outputs of systems which adopt this approach is affected by the number of rules used and their order.

The main obstacle facing the researchers in this field to build an optimal system is the differences in the number of alphabets used in these different languages, and therefore the differences in the phonetics and, therefore, not all letters in different languages have one to one letter correspondence with other languages.

In this study rule-based approach is adopted to evaluate the Romanization system using Back Romanization method, so each Arabic proper noun is transliterated into English, and then the outputted English proper noun is transliterated back into Arabic. The transliteration of any Arabic proper noun is considered a success, if the second transliteration from English into Arabic yields the same source Arabic proper noun inputted to the system. Investigations conducted by the researchers reveal the inability of the used system to transliterate back (i.e. Arabize) the Arabic letter ('ayn, " ξ ") from Romanized proper nouns. The merits of this new method "Back Romanization" are limited to evaluation only, but in its capability to discover the flaws in different Romanization Systems.

We plan to apply this method on a larger dataset that includes geographical names, beside using new methods.

References

- [1] Wikipedia®, List of countries where Arabic is an official language-Wikipedia, the free encyclopedia, Available at: http://en.wikipedia.org/wiki/List_of_countries_w here_Arabic_is_an_official_language, Last visited August 18, 2015.
- [2] Wikipedia®, Arabic language-Wikipedia, the free encyclopedia, Available at: http://en. wikipedia.org/wiki/Arabic- language, Last visited August 18, 2015.

- [3] UNESCO®, World Arabic Language Day, Available at: http://www.unesco.org/new/en/unesco/events/pri zes-and-celebrations/celebrations/internationaldays/world-arabic-language-day/, Last visited August 18, 2015.
- [4] Bateson M. C., "Arabic Language Handbook," Georgetown University Press, 2003.
- [5] Wikipedia®, Arabic alphabet-Wikipedia, the free encyclopedia, Available at: https://en.wikipedia.org/wiki/Arabic_alphabet, Last visited August 18, 2015.
- [6] Habash N., Soudi A. and Buckwalter T., "On Arabic Transliteration," Arabic Computational Morphology: Knowledge based and Empirical Methods, pp. 15-22, 2007.
- [7] Wikipedia®, Buckwalter transliteration-Wikipedia, the free encyclopedia, Available at: http://en.wikipedia.org/wiki/Buckwalter_translit eration, Last visited August 18, 2015.
- [8] Wikipedia®, Latin alphabet-Wikipedia, the free encyclopedia, Available at: http://en.wikipedia.org/wiki/Latin_alphabet, Last visited August 18, 2015.
- [9] Alghamdi M., "Algorithms for Romanizing Arabic Names," Journal of King Saud University: Computer Sciences and Information, (17), pp. 1-27, 2005. (in Arabic).
- [10] Jack H., "Automatic Romanizer of Arabic Names," 2007, Available at: http://www.kanji.org/cjk/arabic/aran.htm, Last visited August 18, 2015.
- [11] Al-Shalabi R., Kanaan G., Al-Sarayreh B., Khanfar K., Al-Ghonmein A., Talhouni H., and Al-Azazmeh S., "Proper Noun Extracting Algorithm for Arabic Language," in Proceedings of the International Conference on Information Technology to Celebrate S. Charmonman's 72nd Birthday, Thailand, 2009.
- [12] Noeman S., "Transliteration using phrase based SMT approach on substrings," in Proceedings of MEDAR 2009: 2nd International Conference on Arabic Language Resources & Tools, Cairo, Egypt; 22-23 April 2009, pp.199-204.
- [13] Brian W., "Lost in Translation," Guardian (UK), 2002, Available at: http://bahailibrary.com/whitaker_lost_in_translation, Last visited August 18, 2015.
- [14] Arabic words and the Roman alphabet, Al-Bab, Available at: http://www.albab.com/arab/language/roman1.htm, Last visited August 18, 2015.
- [15] Romanization System For Arabic: BGN/PCGN 1956, Available at: http://www.adegn.org/PDF/romanization%20sys tem%20ar/1956/1956.pdf, Last visited December 15, 2014.

- [16] Romanization System: IGN 1963 & BGN/PCGN1956, Available at: http://www.adegn.org/Romanization1%20En.htm, Last visited December 15, 2014.
- [17] Romanization System For Arabic-BGN/PCGN 1956 System, Available at: http://geonames.nga.mil/gns/html/, Last visited August 18, 2015.
- [18] Arabic, Report On the Current Status Of United Nations Romanization Systems for Geographical Names, Compiled by the UNGEGN Working Group on Romanization Systems, Version 2.2, January 2003, Available at: http://www.eki.ee/wgrs/rom1_ar.htm, Last visited August 18, 2015.
- [19] UN®, Report On the Current Status Of United Nations Romanization Systems for Geographical Names, Eighth United Nations Conference on The Standardization of Geographical Names, Berlin, 27, August-September 5 2002, Available at:

http://unstats.un.org/unsd/geoinfo/ungegn/docs/8t h-uncsgn-

docs/crp/8th_UNCSGN_econf.94_crp.81.pdf, Last visited August 18, 2015.

- [20] UN®, Updates To the Report On the Current Status Of United Nations Romanization Systems For Geographical Names, New York, 20-29 April 2004. Item No. 11 of the Provisional Agenda, Available at: http://unstats.un.org/unsd/geoinfo/ungegn/docs/22 -GEGN-Docs/wp/gegn22wp48.pdf, Last visited August 18, 2015.
- [21] Romanization System: Beirut 1971, Available at: http://www.adegn.org/Romanization2%20En.htm, Last visited December 15, 2014.
- [22] Arabic, United Nations Romanization Systems for Geographical Names. Preliminary Report on Their Current Status, Compiled by the UNGEGN Working Group on Romanization Systems. Version 1.3. March 2000, Available at: http://www.eki.ee/wgrs/v1_3/rom1_ar.pdf, Last visited August 18, 2015.
- [23] Romanization System: Berlin 2002. Available at: http://www.adegn.org/Romanization3%20En.htm, Last visited December 15, 2014.
- [24] Larkey L. S., AbdulJaleel N., Connell M. "What's in a Name?: Proper Names in Arabic Cross Language Information Retrieval," CIIR Technical Report, IR-278, 2003.
- [25] AbdulJaleel N., Larkey L., "English to Arabic Transliteration for Information Retrieval: A Statistical Approach," CIIR Technical Report IR-261, Department of Computer Science, University of Massachusetts, 2003, Available at: http://ciir.cs.umass.edu/pubfiles/ir-261.pdf, Last visited August 18, 2015.

- [26] Hassan H. & Sorensen J. "An Integrated Approach for Arabic-English Named Entity Translation," in Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages, Ann Arbor, pp. 87–93, 2005.
- [27] Halpern J., "The Challenges and Pitfalls of Arabic Romanization and Arabization," in Proceedings of the Second Workshop on Computational Approaches to Arabic Script-Based Languages. Palo Alto, CA., 2007.
- [28] Halpern J., "Exploiting Lexical Resources for Disambiguating CJK and Arabic Orthographic Variants," in Proceedings of the Sixth International Conference on Language Resources. Morocco, 2008.
- [29] Halpern J., "Lexicon-Driven Approach to the Recognition of Arabic Named Entities," in Proceedings of the Second International Conference on Arabic Language Resources and Tools, 2009, Available at: http://www.elda.org/medarconference/pdf/7.pdf, Last visited August 18, 2015.
- [30] Fattah M. A., Ren F., "English-Arabic propernoun transliteration-pairs creation," JASIST, pp. 1675-1687, 2008.
- [31] Alkharashi I. A., "Person Named Entity Generation and Recognition for Arabic Language," in Proceedings of the second International conference on Arabic Language Resources and Tools. Cairo, Egypt, The MEDAR Consortium, pp. 205-208, 2009.
- [32] David L., "An evaluation of Arabic transliteration methods," A master's paper for M.S degree in L.S degree, School of Information and Library Science, University of North Carolina at Chapel Hill, 2008.
- [33] Lawson D., "An Assessment of Arabic Transliteration Systems," Technical Services Quarterly, Vol. 27, pp. 164–177, 2010.
- [34] Abu Obied H., Nuser M., and Al-Kabi, M. N. A Comparison between Rule & Dictionary Based Romanization of Arabic Names, International Journal of Computer Processing of Oriental Languages, 24(2), 2012, pp. 167-188.
- [35] Arbabi M., Fischthal S. M., Cheng V. C., and Bart E. "Algorithms for Arabic name transliteration," IBM Journal of Research and Development, 38, 2 (March 1994), pp. 183-194, 1994.