Romanizing Arabic Proper Names: Saudi Arabia Experience

Mansour Alghamdi, Ph. D.
King Abdulaziz City for Science and Technology
Riyadh, Saudi Arabia
mghamdi@kacst.edu.sa

Abstract

Although Arabic speakers in the Arab World use Arabic in their daily communications and writing, the transliteration of their Arabic proper names into the Roman/Latin alphabet is necessary nowadays. It is used in many documents and records including: passports, airline tickets, medical, financial, and educational records. However, the used transliteration systems are inconsistent, inappropriate, or unsystematic. These difficulties caused concerns for the security and legal authorities. For the above reasons, Saudi academia and authority held two international symposiums in 2003 and 2006. Experts were asked to participate in the two events to standardize the transliteration of Arabic proper names into English alphabet and the transliteration of foreign proper names into Arabic alphabet. Some of the outcomes of these symposiums are the transliteration table and the algorithms. Based on these outcomes software was developed and now used by different agencies and institutions.

1. Introduction

Romanizing Arabic names is troublesome for officials, employees and name carriers. One reason was the absence of standards. A letter in Arabic may appear in different forms. For example, "محمد" can be transliterated: Muhammad, Mohammed, Mohamed… This creates confusion for officials, employees and name carriers. The first standard was made in Riyadh 2003. It was based on the Arabic writing system, English letters and the existing computer systems that are widely used such as that of the airlines. This paper highlights the efforts made by two eminent institutions in Saudi Arabia namely Naif Arab University for Security Sciences and King Abdulaziz City for Science and Technology in this area.

2. Naif Arab University for Security Sciences (NAUSS) [1]

NAUSS is an Arab academic institution located in Riyadh and affiliated to the Council of the Arab Ministers of Interior within the broader framework of the Arab League. As part of its mission to help modernizing tools and systems for Arab security and immigration personnel, NAUSS held two symposiums on standardizing Arabic name transliteration.

The first symposium was held in 2003 at NAUSS under the name: Standardizing Arabic Names Transliteration: Security Dimensions. Experts in the field were invited to write papers on topics related to the theme of the symposium including: writing Arabic names in the Arabic alphabet, existing Romanization systems of Arabic names, and the problems of the existing Romanization methods. The symposium ended by developing a standardized Romanization table [2].

In 2006, NAUSS held another symposium under the title: Transliteration Between Languages: Romanization of Arabic Proper Names. Experts from different parts of the Arab world were invited to write papers on the theme of the symposium. The titles include: how Arabic names are transliterated into different European languages, transliteration of foreign names into Arabic, methods of transliterating Arabic names into English, and a universally system for transliteration of proper names from one language to another.

Experts and users at the related departments attended the two symposiums where meetings and discussions took place. The symposiums' outcome opens the door for software engineers and phoneticians to develop computer systems that can be used for different applications including health record, security, immigration, travel agencies and educational institutions. Some of the recommendations include:

- English letters to be used only; no other symbols such as diacritics and dash. This will make the transliteration compatible with other important systems such as that of the airlines.
- No difference between capital and small letters; for example not using the capital T for the ط and the small t for ت.
- Parsing should not be included in the transliteration; Muhammad not Muhammadun.
- Compound names to be treated as one; Qamarulzaman not Qamaru Alzaman
- The definite article ال to be transliterated as is without distinguishing between *allamu alshamsiyah* and *allamu alqamariyah*; Alshamsan, Albadrani.

3. King Abdulaziz City for Science and Technology (KACST) [3]

KACST is an independent scientific organization administratively reporting to the Prime Minister of Saudi Arabia and it is located in Riyadh. As part of KACST mission to develop tools and systems that can be utilized by the public and private sectors, it supported a project in 2006 to develop a software system that can Romanize any Arabic name based on the standards that NAUSS come up with in its two symposiums. The team in the project developed the software system in addition to a collection of more than 70,000 Arabic proper names (Figure 1). The system has been available and used since then. Full details about the system are in its final report [4]. However, a summary of its structure is as follows:

The system has four modules. The first is related to Arabic writing including the diacritization system and irregular names such as "طه" and "يسين". This module is important to make the Arabic names ready to be transliterated. The following are short copies of the code:

Arabic names have to be diacritized since an undiacritized name may have more than one pronunciation. For example, حسن *H.s.n* might be حَسَن *Hasan*, a male name, or حُسْن *Husn*, a female name. Moreover, diacritizing Arabic letters has certain rules which need to be strictly applied. Some of these rules are: 1) every letter has to be diacritized that is to be followed by one of the followings:

- *fathah* "أَحْمَد"
- *kasrah* "سِهَام"
- *dhammah* "بُتُول"
- *shaddah+fathah* "سِنِمَّار"
- *shaddah+kasrah* "سُرَيِّع"
- *shaddah+dhammah* "سَلُّوم"
- *sukoun* "عَدْنَان"

The followings are the only exceptions:

- "ا" "سِيمَاء، السَّبْت، لِمَا"
- "آ" "آمَال"
- "و" when it is preceded by *dhammah* and undiacritized "عَلُوش"
- "ي" when it is preceded by *kasrah* and undiacritized "عَلِيم"

- "ى" "رَنَى"
- "ل" when it is preceded by *alif* and followed by *shaddah* "السَّمَاء"

Stage I also takes care of miss-diacritization and some typing habits such as inserting a dash between letters "محـــــــــــــــــــــــــــــــمد" which makes the name looks nice but confuses a computer system.

The second module converts Arabic graphemes into phonetic symbols. Since Arabic letters cover all Arabic sounds, the Arabic alphabet was used to transcribe Arabic sounds. This module covers the aspects such as the *shaddah* and *taa marboutah*. It also calculates the number of syllables in each name to be used in another module.

The third module converts sound symbols into Roman alphabet. The algorithms in this module are based on the standards that came out of the symposiums held at NAUSS. Here each Arabic sound symbol has a rule for its transliteration:

ب	→	b	بَاسِم	Basim
ت	→	t	تَوْفِيق	Tawfeeq
ث	→	th	ثَوْبَان	Thawban
ج	→	j	جُبْرَان	Jubran
ح	→	h	حِزَام	Hizam
خ	→	kh	خَدِيجَة	Khadeejah
د	→	d	دَحَّام	Dahham
ذ	→	th	ذِيب	Theeb
ر	→	r	رَءُوف	Raouf
ز	→	z	زَائِد	Zaed
س	→	s	سَعَيْدَان	Saaydan
ش	→	sh	شَمْس	Shams
ص	→	s	صِنَيْتَان	Sinaytan
ض	→	dh	ضِرَار	Dhirar
ط	→	t	طُوَيْر	Tuwayr
ظ	→	th	ظَاهِر	Thahir
غ	→	gh	غَازِي	Ghazi
ف	→	f	فَايِز	Fayiz
ق	→	q	قَمَرُ الزَّمَان	Qamarulzaman
ك	→	k	كَنْعَان	Kanan
ل	→	l	لُؤَي	Luay
م	→	m	مُحَمَّد	Muhammad
ن	→	n	نَوْف	Nawf
هـ	→	h	هُذَام	Hutham
و	→	w	وَلِيد	Waleed
ي	→	y	يَاسِر	Yasir
ء	→	∅	هَانِئ	Hani
عِـ	→	e	عِمَاد	Emad
عُـ	→	o	العُتَيْبِي	Alotaybi

عـَ/ع	→	a	عَمّار	Ammar
عـْ/	→	e	مَعْن	Maen
ـَ	→	a	سَلْمَان	Salman
ـُ	→	u	غُلَام	Ghulam
ـِ	→	i	نِضَال	Nidhal

The fourth module is a post-processing phase. It includes the capitalization of the name initial and which letters to be used to represent Arabic long vowels according to the number of syllables in the name.

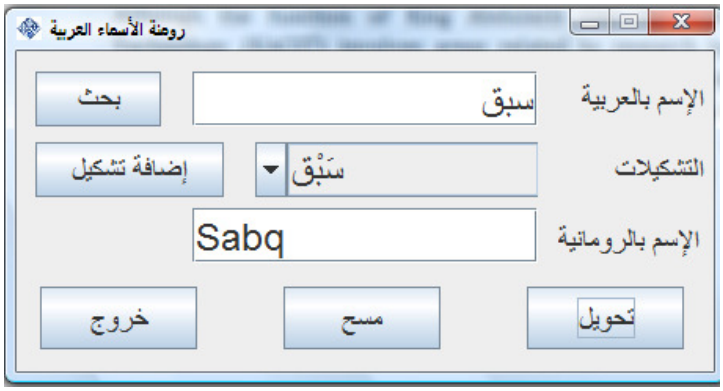

Figure 1. KACST system interface.

4. Conclusion

Many efforts have been made to standardize Arabic name transliteration [5]. However, they were made either for academic research or based on an individual's prospective. As far as the author of this paper knows, there was no serious attempt to standardize Arabic name transliteration for official practice before the first symposium that was held at NAUSS. It was followed by developing a software system that can make the job simple for the people who work on it. The system was developed by KACST and it is available to users who would like to use it.

5. References

[1] Naif Arab University for Security Sciences <http://www.nauss.edu.sa> visited 15 October 2009.

[2] Naif Arab University for Security Sciences, 2003, *The Proceedings of Standardization of Arabic Proper Names Transliteration: Security Dimensions*, Riyadh.

[3] King Abdulaziz City for Science and Technology <http://www.kacst.edu.sa> visited 15 October 2009.

[4] Alghamdi, Mansour, Abdulmalik Alsalman, Abuous Alshamsan, Faysal Almuhanna, Mahmoud Salih, Mansour Alwayili, Khalid Alhuqayl, Salih Alsubai, 2006, Romanization System for Arabic Names (Final Report) KACST.

[5] Salih, Mahmoud, 2003, Romanizing Arabic Names in the Academia. *Standardization of Arabic Proper Names Transliteration: Security Dimensions*, Naif Arab University for Security Sciences, Riyadh, 175-206.