E/CONF.94/INF.77

27 August 2002

Original: English

Eighth United Nations Conference on the
Standardization of Geographical Names
Berlin, 27 August-5 Septernber 2002
Item 16 (a) of the provisional agenda*

WRITING SYSTEMS: ROMANIZATION

Geographic Names Transliteration in GDMS

(Submitted by Saudi Arabia)**

*E/CONF.94/1

** Prepared by Mohaammed Abdulaziz Al-Otaibi, Ministry of Defense & Aviation, General Directorate of Military Suivey GDMS

01-38745 (E) 290501
0135745
02-537646

GEOGRAPHIC NAMES TRANSLITERATION IN GDMS

TRANSLITERATION OF ARABIC GEOGRAPHIC NAMES INTO LATIN RND VICE VERSA

1 INTRUDUCTION

The General Directorate of Military Survey (GDMS) of the Ministry of Defense and Aviation of Saudi Arabia has written software program to enable the transliteration of geographic names from Arabic to Latin. The software allows - as much as possible – going in the reverse direction, i.e. from Latin to Arabic. In tbis paper, we are trying to shed some light on the main aspects and important concerns of this experience.

2 TERMS AND DEFINITIONS[1]

2.1 CHARACTER

A character is an element of a system of writing, whether or not alphabetical, that represents a phoneme, a syllable, the word or even prosodical characteristics of the language by using graphical symbols (letters, diacritical mark, for example ā, ī, ḍ, is therefore a character in the same way as a basic letter).

2.2 TRANSLITERATION

Transliteration is the operation which consists of representing the characters of an entirely alphabetical system of writing by the characters of the conversion alphabet.

2.3 RE-TRANSLITERATION

Re-transliteration is the operation which consists of converting the charactcrs of an alphabet of conversion to those of the alphabet converted.

This operation is the exact opposite of transliteration. It is carried out by applying the d e s of a system of transliteration in reverse order so as to reconstitute the transliterated word to its original form.

2.4 TRANSCRIPTION

Transcription is the operation which consists of representing the characters of a language, whatever the original system of writing, by the phonetic system of letters or signs of the conversion language.

2.5 ROMANIZATION

Romanization is the conversion of non-Latin writing systems to the Latin alphabet. To do that, it is possible to use either transliteration or transcription or a combination of these two methods, according to the nature of the system converted.

3 TRANSLITERATION SYSTEM IN GDMS

The transliteration system used in GDMS to transliterate from Arabic geographic names to Latin is the system that was revised and agreed in 1972. it is based on the system adopted by the Arabic experts at the conference held in Beirut in 1971 with the practica1 amendments carried out and agreed upon by representatives of the Arabic-speaking countries at their conference. It is the system recommended by the United Nations (UN). We will refer to this system later in this documents as *UN transliteration system*, or *UN system* as short.

The rules of UN transliteration system are summarized in the following sections.

3.1 TRANSLITERATION OF THE ARABIC CONSONANTS

Table 1 shows the rules of transliterating Arabic consonants. Note that there is only one form of the letters given in tablc 1, which is isolated form. Rules of Arabic grammar and standards for combining forms in initials, medial, and final positions can be easily found in any Arabic reference and is out of the scope of this paper.

■ *Table 1: Transliteration of the Arabic Consonants.*

NO.	ARABIC LETTER	NAME OF LETTER	TRANSLITERATION INTO ROMAN
1	ء	hamzah	omit (initial) ʾ (medial) ʾ (final)
2	ا	alif	omit
3	ب	bāʾ	b
4	ت	tāʾ	t
5	ث	thāʾ	th
6	ج	Jīm	j
7	ح	ḥāʾ	ḥ
8	خ	khāʾ	kh
9	د	dāl	d
10	ذ	dhāl	dh
11	ر	rāʾ	r
12	ز	zāy	z
13	س	sīn	s

14	ش	chin	ch
15	ص	ṣād	ṣ
16	ض	ḍād	ḍ
17	ط	ṭā′	ṭ
18	ظ	ẓā′	ẓ
19	ع	'ayn	‘
20	غ	ghayn	gh
21	ف	fā′	f
22	ق	qāf	q
23	ك	kāf	k
24	ل	lām	l
25	م	mim	m
26	ن	nün	n
28		waw	w
29	ي	yā′	y

3.2 TRANSLITERATION OF THE ARABIC VOWELS

Usually people omits vowels and diacritical marks from writing, which normally makes it difficult to obtain uniform results in the Romanization of Arabic. The following table summarize the transliteration rules of such letters.

■ *Table 2: Transliteration of the Arabic vowels, diphthongs, and special diacritical marks.*

NO.	ARABIC LETTER	NAME OF LETTER	TRANSLITERATION INTO ROMAN
1	ـَ	fatḥah	a
2	ـِ	kasrah	i

3	ُ	ḍammah	u
4	اَ	fatḥah alif	ā
5	ِي	Kacrah yā′	ī
6	ُو	ḍammah wāw	ū
7	ى	alif maqṣūrah	á
8	ْ	sukūn	omit
9	يْ	fatḥah yā′ sukūn	ay
10	وْ	fatḥah wāw sukūn	aw
11	ً	tanwin fatḥah	a"
12	ٍ	tanwín kasrah	i^n
13	ٌ	tanwin ḍammah	u^n
14	ّ	tashdīd	doubling
15	ا	hamzat al waṣl	′
16	آ	alif maddah	ā (initial) ′ā (medial)

4 DIFFICULTIES IN WRITING TRANSLITERATION SOFTWARE

4.1 TRANSLITERATION RULES

Transliteration rules have to be written in clear and systematic way so it can be processed and understood by aiiy software program.

This need force us to build the system as two main entities. Choosing this model allows us to change some rules and without tlie need to rewrite any piece of code or even any recompilation of tlie software to adopt tlie new changes. So we are able to use "plug-n-play" concept between the rules and the engine.

4.1.1 TRANSLITERATION ENGINE

The first one is tlie transliteration engme. Transliteration engme has all tlie logic and intelligence needed to transliterate a name or a word from one alphabet system to another. The engine has no information about transliteration rules to be used. In other words this engine is transliteration-rules-free.

4.1.2 TRANSLITERATION RULES

This component lias all the rules regarding a specific transliteration system, UN system in our case here.

We have been successful in building the rules-component of both UN system, and for ISO 233 titled "Documentation – Transiiteration of Arabic characters into Latin characters".

4.2 ENCODING SYSTEM

Encoding of characters is always a concern for computer scientists in any Multi-languages system. Fortunately, Unicode encoding system provide a solution. Reader can refer to The Unicode Consortium for more details about Unicode. We used Unicode 3.0 standard.

Mapping of Latin and Arabic letters into Unicode is a straightforward mapping with no ambiguity. But for Latin diacritical marks like Cedilla, and Macron we had to decide on which Unicode encoding to represent them.

The following table shows the chosen mapping of those marks into Unicode. Note that we selected to use what is known as Modifier letters of Unicode to form a combination of letters instead of a pre-coinposed letters.

■ *Table 3: Mapping of Transliteration Latin diacritical marks into Unicode.*

NO.	LATIN MARK	NAME OF MARK	UNICODE VALUE	CHARACTER NAME
1	ˉ	Macron	U+0304	Non-spacing macron
2	´	Acute Accent	U+0301	Non-cpacing acute
3	ʻ	Comrna	U+02BB	Modifier letter turned comrna
4	ʼ	Apostrophe	U+02BC	Modifier letter apostrophe
5	¸	Cedilla	U+0327	Non-spacing cedilla

4.3 REQUIREMMENTS OF INPUT AND OUTPUT

As Transliteration engine expects a well written rules, it also expects a well and clear written input. Arabic writers usually omit vowels from written Arabic as it can deduced by the reader. But this habit can't be practiced over the input of our transliteration engine as it will lead to an output that the user may not expect. This unexpected output is not a malfunction of the system, because the system is working as designed, and expects the right input from the user to get tlie right output.

4.4 REVERSIBILITY

We mean by reversibility the ability to apply re-transliteration as defined above. In our case we transliterate from Arabic alphabet system into Latin alphabet system, in other words, we are doing *Romanization.*

The reverse here would be going from Latin alphabet into Arabic alphabet system, but without using different rules to carry out this task, we can only reverse existing transliteration rules. This is

not a straightfonvard task. The reverse of UN transliteration system will result in some ambiguous rules. For example: an Arabic word that ends with *Taa Marbootah* will be transliterated into a Latin word that ends with *H*. While an Arabic word ends with *Haa* will be transliterated also in a Latin words that ends with *H*.

For example Arabic word "جزيرة" is transliterated to "*Jazīrah*", while Arabic word "ابله" is transliterated into "*Ablah*". So to reverse theses two rules we will get the following two rules that says:

1. an H letter at the end of the word is transliterated into ه
2. an H letter at the end of the word is transliterated into ه

This leads us to say that UN transliteration is not completely reversible. The reason is simply because the systein were built to support a clear task, which is *Romanization.*

5 ADVANTAGES OF OUR SYSTEM IN GDMS

5.1 SOFTWARE FEATURES

We have chosen Visual Basic as a programming language to develop the software, as it is widely used language. The software built over the Windows environment, specifically, Windows NT 4.0, 2000, Me, and XP.

To promote usability of the system, the program were written as a component using a technology known as COM (Component Object Model). Choosing this technology allows an easy deployment and a possible use under different applications that support scripting like MS Word, MS Excel and other Microsoft products. The component can be even deployed to a Server and can be shared by multiple users.

To make it easier, the transliteration rules component are nothing more than a text file written in a predefined format to allow portability of the component.

Our model of implementation of the systems is server-based that serve inultiple users, on a local network, and writing to a database server that contains more than 75,000 geographic names.

5.2 PRODUCTION GAINS

Impleinentation of the systein in GDMS resulted in many advantages to the map production in general and geographical names specifically.

5.2.1 CONSISTENCY

Having the system ensures the consistency of the transliterated naines, where there is no reliance on users' efforts, which results usually in variant output.

5.2.2 TRAINING

The system dropped the training of the names collectors in GDMS drastically. With this systems the names collectors goes through a 2 days training course to be able to start production, without the need to master all the transliteration rules and using them manually.

5.2.3 VERSION CONTROL

The system support tagging the transliteration rules that it uses by a version number. This feature is useful for tracking changes applied to transliteration rules once they occur.

Once the UN geographical names experts update the UN transliteration rules, then the system is capable of applying the new changes while keeping track of the old ones using this tagging feature.

This feature allows the upgrade and downgrade of the rule-set. So the user can transliterate a names that already been translated using previous version of the rule-set using a newer or older version of the nile-set.

6 CONCLUSION

The system has been tested, approved and it is used in geographical names database (more than 75,000 names) and map production at GDMS

We recommend that any suggestions, changes or amendment of the UN transliteration system in the future to take into consideration reversibility (Latin to Arabic) as an important criteria.

We iike our system to be tested and approved by the UN geographical names experts to be a standard transliteration system to be used in all Arabic speaktng countries. That will insure the consistency between the publications from different Arabic speaktng countries.

7 REFERENCEC

1. ISO 233: Documentation – *Transliteration of Arabic characters into Latin characters*, first edition 1984.

2. The Unicode Standard: Worldwide Character Encoding, Version 3.0. The Unicode Consortium.

3. Working Paper No. 34, United Nations group of experts on Geographical Names, Twentieth Session, "United Nations Romanization Systems for Geographical Names, Preliminary Report on their Current Status''

4. BSI, British Standard 4280 : 1968. "Transliteration of Arabic Characters"