

Digits in Arabic Language/Script

Digits Used in our region:

Users who use an Arabic script to write Arabic-based languages (e.g., Arabic, Urdu, Persian ...) are using one or more set of digits in their normal writing without mixing them together in writing numbers. These set are (according to Unicode terminologies):

1. European digits	U+0030 .. U+0039	(0123456789)
2. Arabic-Indic digits	U+0660 .. U0669	(٠١٢٣٤٥٦٧٨٩)
3. Eastern Arabic-Indic digits	U+06F0 .. U+06F9	(۰۱۲۳۴۵۶۷۸۹)

Even in one language community such as the Arabic speaking community, users are using different digits. For example, eastern Arab region (e.g., Egypt, Syria, Sudan, Iraq, all GCC countries, Lebanon, Palestine, Jordan, ...) are mainly using Arabic-Indic digits while the western Arab region (e.g., Libya, Tunisia, Algeria, Morocco, Mauritania, ...) mainly using European digits. But never mixing them together while writing numbers. For example,

1- conference2009	Acceptable: Pure European digits
2- conference٢٠٠٩	Acceptable: Pure Arabic-Indic digits
3- conference٢٠٠٩	Acceptable: Pure Eastern Arabic-Indic digits
4- conference٢٠٠9	Not-Acceptable: Mix between European digits & Arabic-Indic digits
5- conference2٠٠٩	Not-Acceptable: Mix between European digits & Arabic-Indic digits
6- conference٢٠٠٩	Not-Acceptable: Mix between Arabic-Indic digits & Eastern Arabic-Indic digits
7- conference2٠٠٩	Not-Acceptable: Mix between European digits & Eastern Arabic-Indic digits

The Arab Working Group on Arabic Domain Names (AWG-ADN) established by The League of Arab States in 2003 has studied the issue of digits extensively and reached the following recommendations (It was done according to the IDNA 2003):

“Both sets may be supported in the user interface but both must be folded to one set [European] at the preparation of internationalized strings (e.g., "stringprep") phase; i.e. storage of numerals in the zone file is done in ASCII format.”

So we hope this recommendation be honored in the new protocol IDNA 200x, i.e., we need a protocol-level solution to the digit problem.

Number Substitutions in MS Operating Systems

- One important issue with respect to digits is how Microsoft operating systems (Windows XP, 2000, 2003, Vista, ...) treat digits. This means that almost all MS OSs (XP, 2000, Vista, 2003) are storing digits in a unified codes (European digits) and displaying them in the local language setup. According to Microsoft:

“Historically, Windows has supported number substitution by allowing the representation of different cultural shapes for the same digits while keeping the internal storage of these digits unified among different locales, for example numbers are stored

in their well known hexadecimal values, 0x40, 0x41 [European digits], but displayed according to the selected language.

This has allowed applications to process numerical values without the need to convert them from one language to another, for example a user can open an Microsoft Excel spreadsheet in a localized Arabic Windows and see the numbers shaped in Arabic, but open it in a European version of Windows and see European representation of the same numbers. This is also necessary for other symbols such as comma separators and percentage symbol because they usually accompany numbers in the same document.”

Source: [http://msdn.microsoft.com/en-us/library/aa350685\(VS.85\).aspx?PHPSESSID=o1fb21liejulfgrptbmi9dec92#NumberSubstitution](http://msdn.microsoft.com/en-us/library/aa350685(VS.85).aspx?PHPSESSID=o1fb21liejulfgrptbmi9dec92#NumberSubstitution)

- This problem is not for a specific version of MS OS that will be expired but a feature of the operating system behavior that is traditionally implemented by MS in their OSs even the new ones.
- This problem is not for the Saudi Community only but all the Arab region which uses MS OSs.
- Please see Appendix A which shows this problem using Google search. One search is done by typing Pure Arabic-Indic digits while the other is searching by typing Pure European digits but displayed as Arabic-Indic digits.
- Please note that, unfortunately, MS operating systems (XP, 2000, vista ...) are the most widely used OS in our region. Here are some statistics globally and in our region:
http://www.w3schools.com/browsers/browsers_os.asp
http://en.wikipedia.org/wiki/Usage_share_of_desktop_operating_systems
http://www.edunet.tn/webstat/ar/operating_system.htm
<http://www.jamilhamdaoui.net/plugins/log/stats.php?4>
http://www.citc.gov.sa/NR/rdonlyres/2BFE8644-A19C-4CAD-91F8-62BC5ACDC787/0/Internet_Usage_Study_in_KSAIndividualEN.pdf

Digits in Domain Names

- Please note that our discussions with respect the usage of digits are in the scope of domain names, where some restrictions on the size of the character set and the usage is “commonly” imposed for many reasons including security and stability of the domain name system.
- The three sets of digits mean the same (zero to nine) despite their differences in shape.
- With respect to domain names, mixing digit sets (i.e. European and Arabic-Indic) IS NOT applicable and not needed, and hence should be disallowed.
- Users type digits without knowing the internal coding used.

