A Joint Position Statement on Languages Using Arabic Script in Internationalized Domain Names

17 June 2007

Arabic Domain Name Pilot Project   and   IRNIC
Abdulaziz H. Al-Zoman   The ".ir" ccTLD Registry of Iran
Chairman of Steering Committee   Siavash Shahshahani
zoman@isu.net.sa   siavash.shahshahani@irnic.ir

Domain Name System (DNS) is primarily used to locate names on the Internet by translating them into IP addresses. DNS allows only for a limited set of characters (Latin alphabet, digits 0-9 and hyphen) to be used in domain names. With the world-wide spread of the Internet, different nations and language communities are striving hard to get their languages recognized on the DNS system in order to facilitate Internet penetration in their localities.

The Arabic script is the second most widely used alphabetic writing system in the world. It has been adopted for use in a wide variety of languages other than Arabic, including Persian, Urdu, Kurdish, Turkish, Malay, Swahili, … . Such adaptations may feature altered or new characters to represent phonemes that do not appear in Arabic phonology. A new character usually created based on modifying the basic shape of an existing Arabic character, for example, by adding more dots. These additions have meaning to the new language but not to the original language. Hence, supporting languages using Arabic script in domain names necessitate the study of linguistic issues related to each language reaching an accepted character set for each language.

For the purpose of speeding up the deployment of IDN.IDN (i.e., internationalized top level domains), the following principles are agreed on:

- Give the priority to internationalizing ccTLDs. To ensure the system stability, it is recommended to internationalize ccTLDs before internationalizing gTLDs.
- One IDN ccTLD string per one ASCII ccTLD should be applied by the current ASCII ccTLD registry and introduced to the root zone in the initial phase. Recognizing that some countries or regions may have character or language variants, initially only one form of character sets shall be chosen for IDN use by each sponsored registry. Relaxation of this limitation should be revisited after initial deployment of IDN ccTLDs has been experienced.
- The initial choice of IDN ccTLD is made by the current ccTLD Registry with ample comment period for possible objections and reservations by other relevant bodies.
- Register and operate the IDN ccTLDs in the root DNS server in the form of IDNA Punycode.
- With respect to languages using Arabic script, it is recommended that each language be treated separately, i.e., each language (e.g., Arabic, Persian, Urdu, …) have its own accepted character set that should be applied for domain name registration for that language only.
- Supporting multiple languages that use the same script in an IDN-TLD (e.g., IDN gTLDs) demands further studies and investigations that should not prevent the implementation of IDN at the ccTLD level.