

TITLE: Workshop on Internationalized Domain Names for Pakistani Languages

VENUE: National University of Computer and Emerging Sciences, Fast-NU, Lahore, Pakistan.

DATE: April 20, 2008

ATTENDEE LIST

| Name | Designation | Organization |
|-------------------------|--|--|
| Dr. Qasim Bughio | Dean, Faculty of Arts | University of Sindh |
| Dr. Imdad Ali Ismaili | Director, Institute of Information and Communication echnology | University of Sindh |
| Dr. Muhammad Abid | Chairman, Department of Computer Science | University of Peshawar |
| Dr. Salma Shaheen | Director | Pashto Academy |
| Mr. Wahid Buzdar | | Balochi Academy |
| Mr. Zahid Rauf | Incharge, IT Department | Balochistan University of Information and Management Sciences |
| Dr. Muhammad Afzal | Director | Dr. A.Q. Khan Institute of Computer Sciences and Information Technology , Kahuta |
| Mr. Shahzad Ahmad | | BytesForAll |
| Mr. Ashar Nisar | | PKNIC |
| Mr. Hafiz Safwan Chohan | Consultants Urdu Informatics | National Language Authority |
| Dr. Ahsan Wagha | Vice Principal | Pakistan Broadcasting Academy |
| Mr. Arif Aslam Kundi | Director IT | Ministry of IT & Telecom |
| Dr. Shaista Nuzhat | Director | Punjab Institute of Language, Art & Culture |
| Dr. Khaver Zia | Dean, School of Engineering | Beaconhouse National University, Lahore |
| Dr. Sarmad Hussain | Professor & H.O.D | FAST-NU |
| Mr. Shafiq-ur-Rahman | Associate Professor | FAST-NU |
| Mr. Aamir Wali | Lecturer | CRULP, FAST-NU |
| Mr. Atif Gulzar | Senior Research Officer | CRULP, FAST-NU |
| Mr. Agha Ali Raza | Instructor | FAST-NU |
| Mr. Aasim Ali | Phd Scholar | CRULP, FAST-NU |
| Mr. Arfan Mansoor | Research Officer | CRULP, FAST-NU |
| Ms. Seemin Suleri | Regional Research Officer | CRULP, FAST-NU |
| Mr. Inam-ullah | Research Officer | CRULP, FAST-NU |
| Ms. Huda Sarfraz | Senior Research Officer | CRULP, FAST-NU |
| Mr. Zahid Sarfraz | Senior Linguist | CRULP, FAST-NU |
| Mr. Asad Mustafa | Linguist | CRULP, FAST-NU |

REGRETS

| Name | Designation | Organization |
|-------------------|--------------------|--------------------------------------|
| Ms. Jehan Ara | President | Pakistan Software Houses Association |
| Mr. Siddiq Baloch | Director | Balochi Academy |

| | | |
|------------------|---------------------|---------|
| Mr. Bilal Hashmi | Associate Professor | FAST-NU |
|------------------|---------------------|---------|

PROCEEDINGS

The languages which were represented in the workshop are: Balochi, Pashto, Punjabi, Saraiki, Sindhi, and Torwali.

The workshop started with the recitation of the Holy Quran. After that a presentation was delivered by Dr. Sarmad Hussain, Professor & H.O.D of Fast-NU. The main topics of the presentation were as follows:

1. Introduction to Unicode. The points covered in this topic were
 - What is Unicode
 - Why we need Unicode when other encoding systems already exist
 - Universality of the Unicode (standard for all writing systems)
 - Unicode database which specifies character semantics
 - Unicode in use, adopted by industry leaders.
2. Introduction to IDNs. IDNS included the discussion on
 - What is the domain name system (DNS)
 - What are the internationalized domain names (IDNs)
 - Need of the internationalized domain names
 - Internationalized domain names in application (IDNA) and
 - IDNabis protocol
3. Issues and challenges related to Arabic IDNs.
 - Introduction to Arabic Script
 - Characteristics of the Arabic script
 - Positional shapes of different character were discussed
 - Issues in Arabic script were highlighted. Some of which are
 - i. Glyph variants of different character
 - ii. Diacritics issue
 - iii. Bidirectionality issue
 - iv. Normalization issue
 - v. Confusable characters
4. Urdu Language Table for IDNs.

At the end of presentation, Urdu language table for the IDNs was discussed.


The slides of the presentation can be downloaded from (www.crupl.org/idn/download/Presentation.pdf)

After the presentation, there was a discussion session in which all the participants shared their ideas. As expected, there were some conflicts between the participants. There was a suggestion that script should be based on phonemes i.e. should be based (IPA). Although it was good intuition but problem with this is that Internationalized Domain Name (IDNs) work is based on Unicode which is for writing scripts. There was also discussion on whether there should be a single unified script for all languages of Pakistan i.e. there should be one table for Pakistani

languages or each language should have its own separate table. After consensus it was decided that each language should have its separate table.



After discussion session all participants were divided into six groups, one group per language. The task was to decide the character set of each language for the internationalized domain names. As most of participants were linguists, they had some confusions regarding Unicode. For example there are two symbols for 'HEH' (U+0647 and U+06BE) which confused some of participants. To resolve this kind of issues one member from computational field was attached with each group. As a final result of this activity, language tables were finalized for each language (Tables of these languages can be downloaded from www.crup.org/idn/download/LanguageTables.pdf). There were three decisions about the each character of every language: i) Character should be allowed. ii) Character should not be allowed iii) needs discussion (decision pending). Some decisions were taken for all languages after suggestions with each group.

- I. There are three set of digits in Unicode which were considered for all languages, one is from Latin script, and two are from Arabic script (one of Arabic digits and one of Urdu digits). Latin digits are more in use in all the languages under consideration but the local digit block is also used. So the digit block U+06F0 - U+06F9 was included in every language table in addition to Latin numerals
- II. All the languages under discussion require ZERO WIDTH NON JOINER (ZWNJ) for proper display of words. Mostly users are not familiar with use of ZWNJ and use SPACE (U+0020) instead for the same purpose. Since SPACE is not allowed in domain names according to the protocol, this issue is recommended to be handled at application level and SPACE should be replaced with ZWNJ by application before further use.
- III. Decision for Honorifics characters (U+0601 - U+0603) is pending. There was a debate that whether to include these characters or not as they bear some religious significance.  (U+0610) is used with the name of Holy Prophet Hazrat Muhammad (PBUH) very regularly but other honorifics are not mandatory. People were of different views about the inclusion or exclusion of these characters and it was decided that these characters need some further discussion and these will be decided later after consensus.



At the end Mr. Inam-ullah delivered a brief presentation on the languages spoken in northern areas of Pakistan. There are more than 25 distinctive languages spoken in northern areas of Pakistan. The languages which were discussed are: Balti, Bateri, Brushaski, Chiliso, Darmeli, Gawri, Gawro, Gojri, Gwarbati, Indus-Kohistani, kalasha, Kashmiri, Khamveri / Kataveri, Khowar, Khundal Shahi, Oshojo, Palula, Paharri / Pothohari, Shina, Urmuri, Wakhi, Yidgha.

(These slides can be downloaded from www.crupl.org/idn/download/PNL.pdf)



Finally it was decided that there should be one more workshop, which should be an open session. Languages tables will be discussed in that workshop.