GOVERNMENT OF PAKISTAN
MINISTRY OF INFORMATION TECHNOLOGY
(IT & TELECOM DIVISION)
******

No.6-13/2005-IT Islamabad, the 28th April, 2010

Subject: - Minutes of the Meeting of Language Table Sub-Committee on the Development of IDN ccTLD پاکستان.

Enclosed please find herewith minutes of the meeting of Language Table Sub-Committee on the Development of IDN ccTLD (پاکستان.) held on 27th March, 2010 under the Chairmanship of Director (IT), Ministry of IT, Government of Pakistan in the National University of Computer and Emerging Sciences (NUCES), Lahore for information.

Encl: **As Above**

Adnan Saeed
PM NIDU

1. Mr. Abdullah Jan Abid
   Lecturer, AIOU,
   Islamabad

2. Mr. Zia-ur-Rehman
   Lecturer, AIOU,
   Islamabad

3. Mr. Shakeel Ahmed
   Manager Systems, PSEB,
   Islamabad

4. Mr. M. Zaman
   Language Development Consultant, Forum for Language Initiatives,
   Islamabad

5. Mr. Fakhruddin
   Administrator, Frontier Language Institute,
   Peshawar

6. Dr. Khawer Zia
   Dean, Beaconhouse National University,
   Lahore

7. Mr. Taj Joyo
Secretary, Sindhi Language Authority,
Hyderabad

8. Dr. Alamdar Hussain Bhukhari
Director, Saraiki Area Study Centre, B. Z. University,
Multan

9. Miss R. Urooj Bhatti,
Dy. Director Admin, Punjab Institute of Language, Art and Culture,
Lahore

10. Dr. Sarmad Hussain
Professor, NUCES,
Lahore

11. Mr. Inamullah
Research Officer, CRULP,
Lahore

12. Mr. Atif Mirza
Research Officer, CRULP,
Lahore

13. Miss Rabia Sirhinidi
Research Officer, CRULP,
Lahore

14. Miss Huda Sarfraz
Dev. Engineer, CRULP,
Lahore

15. Miss Rahila Parveen
Linguist, CRULP,
Lahore

16. Mr. Asad Mustafa
Linguist, CRULP,
Lahore

CC:

1. P.A to Director (IT)
2. Adnan Saeed, Project Manager (NIDU)
3. M. Abu-ul-Fazal, Project Coordinator (NIDU)
4. Syed Iftikhar Hussain Shah, Analyst (R & D)

Language Table Sub-Committee Meeting for پاکستان. IDN ccTLD

March 27, 2010
National University of Computer and Emerging Sciences, Lahore
Organized by National IT Development & Promotional Unit (NIDU),
Ministry of IT
Government of Pakistan

The first Language Table Sub-Committee Meeting for پاکستان. IDN ccTLD was organized by Ministry of IT on 27th March, 2010 at the National University of Computer and Emerging Sciences (NUCES), Lahore. The language table sub-committee was formed in October 2009 by the Ministry of IT for the development and deployment of Internationalized Domain Names (IDNs) in Pakistan, by its Committee for the Development of پاکستان IDN ccTLD registry. This sub-committee is chartered to finalize a single combined language table for all Pakistani languages based on feedback from all relevant linguistic communities and technical experts. The language table is then going to be submitted along with the IDN ccTLD request for پاکستان. to ICANN by the Ministry.

The meeting on language tables builds on the earlier work done in the context of IDNs for Pakistani languages. This includes former drafts of individual language tables and minutes of the earlier workshop on IDNs organized by MoIT at NUCES in 2009, which have included public participation and feedback, and which have been adopted by the Committee for the Development of پاکستان IDN ccTLD registry of MoIT.

Mr. Arif Aslam Kundi, Director IT, MoIT, chaired the meeting and addressed the audience, putting special emphasis on the role played by technology in reducing the gap between different social segments of the society. Efforts by the Ministry for deployment of necessary infrastructure to enable Internet access in rural areas of the country were highlighted and importance of IDNs was particularly stressed in this regard. Some insight into the ICANN IDN ccTLD and gTLD program was also provided. Finally, he thanked all linguistics and technical experts for coming to contribute to the development of a single language table for the implementation of IDNs in Pakistani languages and acknowledged it as a major milestone in the development of پاکستان. ccTLD registry.

This was followed by a presentation by Dr. Sarmad Hussain, Professor and Head CRULP, NUCES, Lahore. It focused on the work done on developing IDN language tables, address and deploying operational test bed for پاکستان. IDN ccTLD. Information on the establishment of the IDN ccTLD committee and sub-committees for language table, policy and technical test-bed development by MoIT was also presented. The slides of the presentation can be downloaded from http://www.crulp.org/idn/IDN2010.

The following agenda items for the meeting were highlighted.

1. Characters included in the language table for پاکستان. ccTLD.
2. Normalization and extended normalization tables for Pakistani languages
3. Variant table for Pakistani languages

The rest of the proceedings were based on these points, where each was discussed in detail, and finalized by the participants. Summary of relevant discussion and decisions are given below.

I) Language Table

Language table refers to the list of characters that are allowed to be used in IDNs based on language(s) and script(s) used by the community. In the context of Pakistan, it has been decided in the previous meeting on IDNs (minutes available at www.moitt.gov.pk) that a single language table will be developed to support all the 66+ languages spoken in the country. Arabic script is going to be used in IDNs. Based on this decision, opinion was sought from language experts on the characters from each language which are to be included in پاکستان. IDN table.

a. Characters Required for Pakistani Languages

The complete language table listing all characters from all Pakistani languages is given in Appendix-C. This is based on feedback provided by language communities in the April 2008 workshop (which can be accessed at http://www.crulp.org/idn/download/LanguageTables.pdf) and was used as the starting point of reference for the current discussion. Highlighted characters in the table are allowed to be registered in IDNs and will be added on the keyboard as well. Characters that are not highlighted cannot be registered in IDNs, either because they are not required by any Pakistani language or because they can cause potential security risks such as diacritics. The use of diacritics is not allowed at this time but may be allowed with advancements in technology to cater for security problems, as was agreed in the previous meeting. Feedback on the language table by respective language representatives is summarized below, and detailed character sets are given in Annexure G.

Balochi – All characters required for Balochi language are highlighted in the language table. No additions are required.

Brahvi – All characters required for Brahvi language are highlighted in the table. No additions are required.

- **Gawri** – All characters required for Gawri language are highlighted in the table. No additions are required. U+065E ( ) is used in the language but only as a diacritic (e.g. in). As diacritics are currently not allowed in IDNs, therefore, this character cannot be included in the table. This character is in the set of optional aerab in Gawri and therefore not limit expression in the language more than exclusion of other aerabs.

Khowar – All characters required for Khowar language are highlighted in the table. No additions are required.

Pashto – All characters required for Pashto language are highlighted in the table. No additions are required.

Punjabi – All characters required for Punjabi language are highlighted in the table. Only one character ڶ is required by the language community that is not present in the Unicode Standard. *Therefore, it is recommended that the character be encoded in Unicode through a separate process. Once it is included in the Unicode, it will be included in the language table.*

Sariaki – All characters required for Saraiki language are highlighted in the table, except ﺽ which should be added to the table. Only one character ﮒ is required by the language community that is not present in the Unicode Standard. *Therefore, it is recommended that the character be encoded in Unicode through a separate process. Once it is included in the Unicode, it will be included in the language table.*

Sindhi – The character U+0647 (ه) is required by the Sindhi language community for use in the initial position of words. Therefore, it has to be included in the language table. All other characters are highlighted already.

Torwali – All characters required for Torwali language are highlighted in the table. No additions are required.

Urdu – All characters required for Urdu language are highlighted in the table. No additions are required.

b. Combining Marks (Diacritics and Honorifics)

All participants agreed upon the set of combining marks for IDNs in Pakistani languages as given in Appendix-D. These include honorifics (U+0610..U+0614) and diacritics (U+064B..U+065E, U+0670). Thus, the diacritics requirement for all languages is complete.

However, as per agreement reached in the previous meeting, honorifics would not be allowed initially, until technology is updated to make their use secure.

c. ASCII Block

Letter-Digit-Hyphen (LDH) characters from the ASCII block can be included in the language table for Pakistani languages. This raises a few questions like whether these characters can be mixed with local languages' characters or not. Mixing of scripts can occur in the following ways.

1. Mixing of scripts within labels. For e.g., in پاکستان.abcاردو
2. Mixing of scripts across labels. For e.g., in پاکستان.abc
3. No mixing. For e.g., in اردو.پاکستان

One group of participants was of the view that mixing of ASCII characters with Arabic was not required for Pakistani domain names. Allowing registration of ASCII labels will also open up potential for phishing due to confusion created through mixed writing directions of the two scripts (e.g. bب may be confused with بb). In addition, mixed script labels are difficult to type and keyboard switching is required-a process that is not very user-friendly.

Another group of participants, however, supported mixing of ASCII and Arabic characters within and across labels. They were of the view that ASCII characters are being used for abbreviations and save time while typing as compared to complete strings. Also, ASCII characters are used by some Northern languages of Pakistan, therefore these should be allowed. Allowing them would also give more flexibility to the users.

d. Join Controls

It was decided in the previous meeting that Zero Width Non-Joiner (ZWNJ) and Hyphen-Minus (-) will be used to demarcate words and sub-words in Pakistani IDNs, e.g. زبان-اردو. However, Hyphen-Minus should be tested before it is allowed, as it can also cause script mixing related problems: e.g. users may get confused between زبان-اردو and اردو-زبان. This decision is also supported by the language table sub-committee.

Decisions and Recommendations

The following decisions were made with respect to allowable characters in the language table.

1. Based on discussion on table in Annexure C-I, the tables in Annexures C-II, C-III and D are approved, with the following additional recommendations.
2. U+0647 and U+06FB are added to the list of allowed characters in the common language table for Pakistani IDNs.

3. Both ZWNJ and Hyphen-Minus (-) are to be adopted for use in IDNs for Pakistani languages. However, Hyphen-Minus would be tested with applications and user feedback would be taken into account before it can be allowed in domain name registrations to ensure it does not cause user confusion.
4. Confusing Arabic language characters U+0643 (ك) and U+0649 (ى) are not required by any language community, and should be handled through variants (see the section on Variant). They are removed from the language table of required characters.
5. Mixing of scripts is not allowed currently in IDNs implementation. This includes both cases of mixed script labels as discussed in the previous section. However, پاکستان. ccTLD registry may allow registration of ASCII-only labels once the system is stable for local languages' domain names. This refers to only allowing mixing across labels. Much later the registry can allow mixed script labels, if there is a significant need and security issues are addressed. User and application testing should be done before script mixing is allowed at any degree, to completely avoid issues due to confusability.
6. Missing characters □ل for Punjabi and □گ for Siraiki should be proposed to be added in the Unicode Standard by the Ministry.

II) Normalization and Extended Normalization Tables

Sometimes Unicode provides multiple representations for one character in the same script block. A character may be encoded directly, or formed by two or more composing parts. This results in two different Unicode sequences corresponding to the same character. For example آ = U+0627 + U+0653 is the decomposed form of the composite character آ = U+0622. Both look exactly the same, creating a likely user confusion. Unicode provides a normalization process to compose both forms to a single representation, through Normalization Form C (NFC). Annex-E-I provides a complete list of Arabic characters and their corresponding normalized forms as provided by Unicode.

In addition to normalization provided by Unicode, there are some Arabic characters for which Unicode does not provide normalized equivalents. A complete list of such composite characters and their decomposed forms is given in Annex-E-II. These additional normalizations should also be done to avoid user confusions at the registry level.

Decisions and Recommendations

The following decisions are made.

1. All participants agreed to adopt the existing normalization and the extended normalization tables given in Annexure E-I and E-II, while the latter will be implemented at the registry.

2. In addition, a proposal needs to be sent to the Unicode consortium to incorporate as much of the extended normalization in the regular normalization process, e.g. for خ.

III) Variant Table

In addition to the pre-composed characters, there are some visually confusable characters within the Arabic script. Mapping has to be defined at an appropriate level to maintain equivalence of such characters and reduce user confusion. Each character can take one of the four visual forms depending upon its context in a word. These are (i) Initial, (ii) Medial, (iii) Final and (iv) Isolated forms. Characters that look similar evening any *one* of the four forms are said to be visually confusable. For e.g. Arabic script provides two versions for the letter Kaf; an Arabic version ك (U+0643) and a Persian version ک (U+06A9). Both look exactly the same in their initial and medial positions. Finally, mapping may also be required at the language level, as two characters may be distinct at the script level, but may be confusingly similar at the language level. An example is that of the Arabic Kaf ک (U+06A9) and the Arabic Swash Kaf ڪ (U+06AA). In Sindhi, the two are distinct letters but in other Pakistani languages the latter would be considered stylistic variation of the former character. Decisions have already been taken to map these characters in the earlier meetings. The current work was to finalize the list for which mapping has to be done.

Variant characters are further divided into three categories based on what makes them confusingly similar to each other. The following sections provide the decisions taken by all language communities regarding specific cases of confusable characters.

a. Variants Based on Shape Similarity

Shape based variants are characters that form exactly the same shapes in one or more of the four positions discussed above. This category also includes characters that do not look exactly similar but may be considered stylistic variations of each other by a language community in Pakistan. For example consider two strings بٹجو and بٹجو with characters ٹ and ٹ respectively. Both appear exactly the same in medial positions. Though they are not used concurrently in a single language, in the context of Pakistani languages being used simultaneously, such examples may arise. Further, the labels ﭘﯽ and ﭘﯽ with characters ی and ی in final positions may look stylistic variations to non-Pashto speakers.

Discussion

All participants agreed on the fact that shape-based variants (whether exact or similar looking) may cause confusion among the users and create a potential for phishing in IDNs. ASCII digits are also used in a similar fashion as other local language digits. They are not considered different by users. So they should also be mapped on to each other.

Therefore, these should be mapped onto each other during registration and lookup, as given in Table 1 below.

Table 1. List of Variants for Pakistani Languages Based on Shape Similarity

Character	Unicode	Character	Unicode	Mapping Decision
ي	064A	ی	06CC	Map in initial and medial positions
ی	06CD	ی	06CC	Map in final and isolated positions
ہ	0647	ھ	06BE	Map in all four positions
ہ	0647	ہ	06C1	Map in all four positions
ک	06A9	ک	06AA	Map in all four positions
ٹ	0679	ٹ	06BB	Map in initial and medial positions
۰	06F0	0	0030	Yes
۱	06F1	1	0031	Yes
۲	06F2	2	0032	Yes
۳	06F3	3	0033	Yes
۴	06F4	4	0034	Yes
۵	06F5	5	0035	Yes
۶	06F6	6	0036	Yes
۷	06F7	7	0037	Yes
۸	06F8	8	0038	Yes
۹	06F9	9	0039	Yes

Decisions and Recommendations

Based on the above given feedback provided by representatives of all language communities, the following was decided:

1. The characters that have similar shapes in any position should be mapped to each other and any two labels only differing in such a pair of variant characters should be treated as equivalent. For example پاکستان (with U+06A9) and پاکستان (with U+0643) will be treated as the same labels.
2. Mapping should be done as per Table 1 above.

b. Variants Based on Confusability with Arabic Language Characters

Some characters in the language table for Pakistani languages have shape variants in the Arabic language character set. For example the Arabic letter Kaf (ك) exactly looks similar to Arabic letter Keheh (ک) in initial and medial positions. Pakistani languages do not require the Arabic version of Kaf and hence it is disallowed in IDNs for Pakistani languages. However, Arabic keyboards support Arabic letter Kaf (U+0643) and not the other version. This leads to a potential problem where users using Arabic keyboards will not be able to access Pakistani websites using Arabic letter Keheh (U+06A9) since the keyboard does not support it (e.g. the expat Pakistani community in the Middle East). Therefore, ك needs to be mapped to ک at resolution time to avoid DNS lookup failures, though websites with these letters should not be registered.

Table 2. List of Variants Based on Confusability with Arabic Language Characters

Character	Unicode	Character	Unicode	Mapping Decision
ة	0629	ۃ	06C3	Map during resolution
ک	06A9	ك	0643	Map during resolution
ک	06AA	ك	0643	Map during resolution
ی	06CC	ى	0649	Map during resolution
ي	064A	ى	0649	Map during resolution
ۍ	06CD	ى	0649	Map during resolution
۰	06F0	◌	0660	Map during resolution
۱	06F1	◌	0661	Map during resolution
۲	06F2	◌	0662	Map during resolution
۳	06F3	◌	0663	Map during resolution
۴	06F4	◌	0664	Map during resolution
۵	06F5	◌	0665	Map during resolution
۶	06F6	◌	0666	Map during

				resolution
۷	06F7	◎	0667	Map during resolution
۸	06F8	◎	0668	Map during resolution
۹	06F9	◎	0669	Map during resolution

Discussion

All participants were of the view that Arabic language characters should not be allowed in Pakistani IDNs. However, due to the possibility of occurrence of some characters in IDNs at the resolution time, it is needed that these should be mapped to the ones present in Pakistani language table.

Decisions and Recommendations

Based on general consensus the following has been recommended:

1. Arabic characters that are visually confusable with any characters from Pakistani language table should be mapped during domain name lookup.
2. The complete list of characters is given in Table 2 above.

c. Variants Based on Orientation of Dots

These variants comprise of characters that look confusing because of the direction of dots on a base character. For example U+0683 (ڃ) and U+0684 (ڄ) have the same number of dots, only their direction is different. This may cause confusion among users who are not familiar with the languages these characters belong to. Thus they may also be considered variants.

Table 3. List of Variant Characters Based on Dot Orientation

Character	Unicode	Character	Unicode	Mapping Decision
ڃ	0683	ڄ	0684	No. But reserve label with confusable character until further analysis
ٹ	062A	ٺ	067A	No. But reserve label with confusable character until further analysis

ٹ	062B	ٹ	067D	No. But reserve label with confusable character until further analysis
ي	064A	ی	06D0	No. But reserve label with confusable character until further analysis

Discussion

Though there was a general agreement among participants to map shape confusable characters, the discussion on dot orientation did not conclude. It was not clear if variant characters based on dot orientation were confusable. Some participants found these pairs of characters confusing (for e.g., ٹ and ٹ), as they exhibited a minimal difference in the positioning of dots. This especially became confusing in the browser's URL bar where font sizes were very small (see Figure 1 below). However, other participants claimed that these were not confusable characters and should not be mapped.

Figure 1. Addresses Containing Dot Orientation Based Variant Characters (ٹ and ٹ)

Decisions and Recommendations

Owing to difference of opinion with the various language communities, the following has been recommended.

1. For variants based on dot orientation, one label would be registered at a time and the other would be reserved (neither mapped nor separately delegated).
2. A user survey and IDN deployment experience is needed for further decision.
3. Once the decision is done, then the reserved label can be released for registration in future to the same or different registrant.

This allows flexibility to either allow separate delegations or mapping of the existing delegations in future, based on the experience gained in the management of registry.

Conclusion and Recommendations

The participants agreed on the contents of language table, normalization and variant analysis, as has been described. These have been tabulated. However, as IDNs are a new and untested technology, a conservative approach has been recommended, in which minimal character set is used for initial deployment. As we and others around the World gain more experience in IDNs and become more aware of security threats and usability issues, the table may be expanded in the future. Appendix-F provides the language table listing characters and their variants to be submitted to IANA for پاکستان. IDN ccTLD currently. This table may need to be updated if characters needed in the future are to be submitted at this time to IANA with the Fast Track application.

Annex-A: Language Table Sub-Committee Meeting on Internationalized Domain Names (IDNs) for پاکستان. IDN ccTLD

Program for 27th March, 2010

Time	Agenda
08:30 – 09:00	Registration and Tea
09:00 – 09:15	Welcome address (by Ministry of IT)
09:15 – 10:00	Introduction to IDNs and Meeting Work Plan (by Dr. Sarmad Hussain)
10:00 – 12:30	(i) Character-by-character analysis of Language table
12:30 – 01:15	(ii) Normalization table (NFC + Extended Normalization)
01:15 – 02:00	Lunch and prayer break
02:00 – 04:30	(iii) Variant table
04:30 – 04:45	Tea

Annex-B: List of Attendees

Language Table Sub-Committee Meeting on Internationalized Domain Names (IDNs) for پاکستان. IDN ccTLD

27th March, 2010

S#	Name	Designation	Organization
1	ADNAN SAEED	Project Manager	M/o IT
2	ARIF ASLAM Kundi	IT Expert	———
3	ABDULLAH JAMALI	Lecturer	AIOU Islamabad
4	Zia ur Rehman Baloch	Lecturer	AIOU Islamabad
5	Shakeel Ahmed	Manager Systems	PSEB
6	M Zamaan Sajjad	Language Dev Consultant	Forum for Language Initiative
7	FAKHRUDDIN	Administrator	FLI
8	INAM ULLAH	Research Officer	CRULP
9	ATIF MIRZA	Research officer	CRULP
10	Rahat Sirhindi	Research Officer	CRULP
11	Hudaibahoo	Dep. Engineer	CRULP
12	Rahila Parveen	Linguist	CRULP
13	Asad Mustafa	S. Linguist	CRULP
14	DR KHAVER ZIA	DEAN, [redacted]	BNU
15	Taj Joyo	Secretary	Sindhi Language Authority Hyd
16	Dr Alamdar Hussain Bukhari	Director SASC	Siraiki Area Study Centre BZ Univ.
17	M Abu ul-Fazl	Project Coordinator	M/o IT
18	Syed Zulfikhar H Shah	Analyst	M/o IT
19	R. UROOJ BHATTI	Dy. Dir Admn	PILAC
20	Farzand Ali	Asst. [illegible]	Gulf
21	M. Junaid	Assistant Manager Admin & HR	CRULP
22	Mustafa	Office boy	CRULP
23	Dil Mohd Ali	A—O	CRULP

Annex-C: Language Table for Pakistani Languages

C-I: Draft of Language Table Circulated for Discussion (based on earlier meeting)

Complete Language Table Listing
Characters to be Allowed for use
in IDNs for Pakistani Languages

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F	075	076	077
0	0600	0610		0630	0640	0650	0660	0670	0680	0690	06A0	06B0	06C0	06D0	06E0	06F0	0750	0760	0770
1	0601	0611	0621	0631	0641	0651	0661	0671	0681	0691	06A1	06B1	06C1	06D1	06E1	06F1	0751	0761	0771
2	0602	0612	0622	0632	0642	0652	0662	0672	0682	0692	06A2	06B2	06C2	06D2	06E2	06F2	0752	0762	0772
3	0603	0613	0623	0633	0643	0653	0663	0673	0683	0693	06A3	06B3	06C3	06D3	06E3	06F3	0753	0763	0773
4		0614	0624	0634	0644	0654	0664	0674	0684	0694	06A4	06B4	06C4	06D4	06E4	06F4	0754	0764	0774
5		0615	0625	0635	0645	0655	0665	0675	0685	0695	06A5	06B5	06C5	06D5	06E5	06F5	0755	0765	0775
6	0606	0616	0626	0636	0646	0656	0666	0676	0686	0696	06A6	06B6	06C6	06D6	06E6	06F6	0756	0766	0776
7	0607	0617	0627	0637	0647	0657	0667	0677	0687	0697	06A7	06B7	06C7	06D7	06E7	06F7	0757	0767	0777
8	0608	0618	0628	0638	0648	0658	0668	0678	0688	0698	06A8	06B8	06C8	06D8	06E8	06F8	0758	0768	0778
9	0609	0619	0629	0639	0649	0659	0669	0679	0689	0699	06A9	06B9	06C9	06D9	06E9	06F9	0759	0769	0779
A	060A	061A	062A	063A	064A	065A	066A	067A	068A	069A	06AA	06BA	06CA	06DA	06EA	06FA	075A	076A	077A
B	060B	061B	062B	063B	064B	065B	066B	067B	068B	069B	06AB	06BB	06CB	06DB	06EB	06FB	075B	076B	077B
C	060C		062C	063C	064C	065C	066C	067C	068C	069C	06AC	06BC	06CC	06DC	06EC	06FC	075C	076C	077C
D	060D		062D	063D	064D	065D	066D	067D	068D	069D	06AD	06BD	06CD	06DD	06ED	06FD	075D	076D	077D
E	060E	061E	062E	063E	064E	065E	066E	067E	068E	069E	06AE	06BE	06CE	06DE	06EE	06FE	075E	076E	077E
F	060F	061F	062F	063F	064F		066F	067F	068F	069F	06AF	06BF	06CF	06DF	06EF	06FF	075F	076F	077F

C-II: Final Language Table based on Sub-Committee Meeting (changes marked in Red)

Complete Language Table Listing
Characters to be Allowed for use
in IDNs for Pakistani Languages

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F	075	076	077
0	0600	0610		0630	0640	0650	0660	0670	0680	0690	06A0	06B0	06C0	06D0	06E0	06F0	0750	0760	0770
1	0601	0611	0621	0631	0641	0651	0661	0671	0681	0691	06A1	06B1	06C1	06D1	06E1	06F1	0751	0761	0771
2	0602	0612	0622	0632	0642	0652	0662	0672	0682	0692	06A2	06B2	06C2	06D2	06E2	06F2	0752	0762	0772
3	0603	0613	0623	0633	0643	0653	0663	0673	0683	0693	06A3	06B3	06C3	06D3	06E3	06F3	0753	0763	0773
4		0614	0624	0634	0644	0654	0664	0674	0684	0694	06A4	06B4	06C4	06D4	06E4	06F4	0754	0764	0774
5		0615	0625	0635	0645	0655	0665	0675	0685	0695	06A5	06B5	06C5	06D5	06E5	06F5	0755	0765	0775
6	0606	0616	0626	0636	0646	0656	0666	0676	0686	0696	06A6	06B6	06C6	06D6	06E6	06F6	0756	0766	0776
7	0607	0617	0627	0637	0647	0657	0667	0677	0687	0697	06A7	06B7	06C7	06D7	06E7	06F7	0757	0767	0777
8	0608	0618	0628	0638	0648	0658	0668	0678	0688	0698	06A8	06B8	06C8	06D8	06E8	06F8	0758	0768	0778
9	0609	0619	0629	0639	0649	0659	0669	0679	0689	0699	06A9	06B9	06C9	06D9	06E9	06F9	0759	0769	0779
A	060A	061A	062A	063A	064A	065A	066A	067A	068A	069A	06AA	06BA	06CA	06DA	06EA	06FA	075A	076A	077A
B	060B	061B	062B	063B	064B	065B	066B	067B	068B	069B	06AB	06BB	06CB	06DB	06EB	06FB	075B	076B	077B
C	060C		062C	063C	064C	065C	066C	067C	068C	069C	06AC	06BC	06CC	06DC	06EC	06FC	075C	076C	077C
D	060D		062D	063D	064D	065D	066D	067D	068D	069D	06AD	06BD	06CD	06DD	06ED	06FD	075D	076D	077D
E	060E	061E	062E	063E	064E	065E	066E	067E	068E	069E	06AE	06BE	06CE	06DE	06EE	06FE	075E	076E	077E
F	060F	061F	062F	063F	064F		066F	067F	068F	069F	06AF	06BF	06CF	06DF	06EF	06FF	075F	076F	077F

C-III: ASCII and Join Control Characters Included in the Final Language Table for پاکستان. IDN ccTLD

ASCII Characters and Join Control Characters Allowed to be used in IDNs for Pakistani Languages

000	001	002	003	004	005	006	007	200	201	
NUL 0000	DLE 0010	SP 0020	0 0030	@ 0040	P 0050	` 0060	p 0070	NQ SP 2000	- 2010	0
SOH 0001	DC1 0011	! 0021	1 0031	A 0041	Q 0051	a 0061	q 0071	MQ SP 2001	NB - 2011	1
STX 0002	DC2 0012	" 0022	2 0032	B 0042	R 0052	b 0062	r 0072	EN SP 2002	– 2012	2
ETX 0003	DC3 0013	# 0023	3 0033	C 0043	S 0053	c 0063	s 0073	EM SP 2003	— 2013	3
EOT 0004	DC4 0014	$ 0024	4 0034	D 0044	T 0054	d 0064	t 0074	3/M SP 2004	―― 2014	4
ENQ 0005	NAK 0015	% 0025	5 0035	E 0045	U 0055	e 0065	u 0075	4/M SP 2005	―― 2015	5
ACK 0006	SYN 0016	& 0026	6 0036	F 0046	V 0056	f 0066	v 0076	6/M SP 2006	‖ 2016	6
BEL 0007	ETB 0017	' 0027	7 0037	G 0047	W 0057	g 0067	w 0077	F SP 2007	‗ 2017	7
BS 0008	CAN 0018	(0028	8 0038	H 0048	X 0058	h 0068	x 0078	P SP 2008	‘ 2018	8
HT 0009	EM 0019	) 0029	9 0039	I 0049	Y 0059	i 0069	y 0079	TH SP 2009	’ 2019	9
LF 000A	SUB 001A	* 002A	: 003A	J 004A	Z 005A	j 006A	z 007A	H SP 200A	‚ 201A	A
VT 000B	ESC 001B	+ 002B	; 003B	K 004B	[005B	k 006B	{ 007B	ZW SP 200B	‛ 201B	B
FF 000C	FS 001C	, 002C	< 003C	L 004C	\ 005C	l 006C	| 007C	ZW NJ 200C	“ 201C	C
CR 000D	GS 001D	- 002D	= 003D	M 004D	] 005D	m 006D	} 007D	ZW J 200D	” 201D	D
SO 000E	RS 001E	. 002E	> 003E	N 004E	^ 005E	n 006E	~ 007E	LRM 200E	„ 201E	E
SI 000F	US 001F	/ 002F	? 003F	O 004F	_ 005F	o 006F	DEL 007F	RLM 200F	‟ 201F	F

Annex-D: Combining Characters Table

Combining Characters (Diacritics and Honorifics) Currently NOT Allowed in IDNs for Pakistani Languages

May be allowed at a later stage after formal consideration

	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F	075	076	077
0	0600	0610		0630	0640	0650	0660	0670	0680	0690	06A0	06B0	06C0	06D0	06E0	06F0	0750	0760	0770
1	0601	0611	0621	0631	0641	0651	0661	0671	0681	0691	06A1	06B1	06C1	06D1	06E1	06F1	0751	0761	0771
2	0602	0612	0622	0632	0642	0652	0662	0672	0682	0692	06A2	06B2	06C2	06D2	06E2	06F2	0752	0762	0772
3	0603	0613	0623	0633	0643	0653	0663	0673	0683	0693	06A3	06B3	06C3	06D3	06E3	06F3	0753	0763	0773
4		0614	0624	0634	0644	0654	0664	0674	0684	0694	06A4	06B4	06C4	06D4	06E4	06F4	0754	0764	0774
5		0615	0625	0635	0645	0655	0665	0675	0685	0695	06A5	06B5	06C5	06D5	06E5	06F5	0755	0765	0775
6	0606	0616	0626	0636	0646	0656	0666	0676	0686	0696	06A6	06B6	06C6	06D6	06E6	06F6	0756	0766	0776
7	0607	0617	0627	0637	0647	0657	0667	0677	0687	0697	06A7	06B7	06C7	06D7	06E7	06F7	0757	0767	0777
8	0608	0618	0628	0638	0648	0658	0668	0678	0688	0698	06A8	06B8	06C8	06D8	06E8	06F8	0758	0768	0778
9	0609	0619	0629	0639	0649	0659	0669	0679	0689	0699	06A9	06B9	06C9	06D9	06E9	06F9	0759	0769	0779
A	060A	061A	062A	063A	064A	065A	066A	067A	068A	069A	06AA	06BA	06CA	06DA	06EA	06FA	075A	076A	077A
B	060B	061B	062B	063B	064B	065B	066B	067B	068B	069B	06AB	06BB	06CB	06DB	06EB	06FB	075B	076B	077B
C	060C		062C	063C	064C	065C	066C	067C	068C	069C	06AC	06BC	06CC	06DC	06EC	06FC	075C	076C	077C
D	060D		062D	063D	064D	065D	066D	067D	068D	069D	06AD	06BD	06CD	06DD	06ED	06FD	075D	076D	077D
E	060E	061E	062E	063E	064E	065E	066E	067E	068E	069E	06AE	06BE	06CE	06DE	06EE	06FE	075E	076E	077E
F	060F	061F	062F	063F	064F		066F	067F	068F	069F	06AF	06BF	06CF	06DF	06EF	06FF	075F	076F	077F

Annex-E: Normalization Tables

E-I: Normalization Table (NFC provided by Unicode)

Desired Combined Form (Unicode)	Decomposed Form		Description
	Character 1	Character 2	
آ 0622	ا 0627	ٓ 0653	ARABIC LETTER ALEF WITH MADDA ABOVE
أ 0623	ا 0627	ٔ 0654	ARABIC LETTER ALEF WITH HAMZA ABOVE
ؤ 0624	و 0648	ٔ 0654	ARABIC LETTER WAW WITH HAMZA ABOVE
إ 0625	ا 0627	ٕ 0655	ARABIC LETTER ALEF WITH HAMZA BELOW
ئ 0626	ي 064A	ٔ 0654	ARABIC LETTER YEH WITH HAMZA ABOVE
ۂ 06C2	ہ 06C1	ٔ 0654	ARABIC LETTER HEH GOAL WITH HAMZA ABOVE
ۓ 06D3	ے 06D2	ٔ 0654	ARABIC LETTER YEH BARREE WITH HAMZA ABOVE

E-II: Extended Normalization Table (not provided by Unicode)

Character (Unicode)	Decomposed Form		Description
	Character 1	Character 2	
ئ 0626	ى 0649	ٔ 0654	ARABIC LETTER YEH WITH HAMZA ABOVE
ئ 0626	ی 06CC	ٔ 0654	ARABIC LETTER YEH WITH HAMZA ABOVE
ۂ 06C2	ہ 0647	ٔ 0654	ARABIC LETTER HEH GOAL WITH HAMZA ABOVE
ځ 0681	ح 062D	ٔ 0654	ARABIC LETTER HAH WITH HAMZA ABOVE

Annex-F: Variant Table For Submission to IANA Based on Sub-Committee Meeting

TLD: PK
Language Tag: *-PK
Language Description: Pakistani languages
Version: 1.0
Effective Date:

Registry:
Contact:
Address:
Telephone:
Website:

This document provides a description of the IDN (Internationalized Domain Names) Language Table to be used by the Pakistan IDN ccTLD registry for the registration of Pakistani languages. These are based on the recommendation of the پاکستان. IDN ccTLD Language Table Sub-Committee.

#Characters from Unicode Arabic Table (0600-06FF)
U+0621
U+0622
U+0623
U+0624
U+0625
U+0626
U+0627
U+0628
U+062A;U+067A (T,-)
U+062B;U+067D (T,-)
U+062C
U+062D
U+062E
U+062F
U+0630
U+0631
U+0632
U+0633
U+0634
U+0635
U+0636
U+0637
U+0638
U+0639
U+063A
U+0641
U+0642
U+0644
U+0645
U+0646
U+0647;U+06BE (E,InMFIs);U+06C1 (E,InMFIs)
U+0648

```
U+064A;U+06CC (E,InM);U+06D0 (T,-)
U+0679;U+06BB (T,InM)
U+067A;U+062A (T,-)
U+067B
U+067C
U+067D;U+062B (T,-)
U+067E
U+067F
U+0680
U+0681
U+0683;U+0684 (T,-)
U+0684;U+0683 (T,-)
U+0685
U+0686
U+0687
U+0688
U+0689
U+068A
U+068B
U+068C
U+068D
U+068F
U+0691
U+0693
U+0696
U+0698
U+0699
U+069A
U+06A6
U+06A9;U+06AA (T,InMFIs)
U+06AA;U+06A9 (T,InMFIs)
U+06AB
U+06AF
U+06B1
U+06B3
U+06B7
U+06BA
U+06BB;U+0679 (T,InM)
U+06BC
U+06BE;U+0647 (E,InMFIs); 06C1 (E,InMFIs)
U+06C1;U+0647 (E,InMFIs); 06BE (E,InMFIs)
U+06C2
U+06C3
U+06CC;U+064A (E,InM);U+06CD (T,FIs)
U+06CD;U+06CC (T,FIs)
U+06D0;U+064A (T,-)
U+06D2
U+06D3
U+06F0;U+0030 (T,-)
```

```
U+06F1;U+0031 (T,-)
U+06F2;U+0032 (T,-)
U+06F3;U+0033 (T,-)
U+06F4;U+0034 (T,-)
U+06F5;U+0035 (T,-)
U+06F6;U+0036 (T,-)
U+06F7;U+0037 (T,-)
U+06F8;U+0038 (T,-)
U+06F9;U+0039 (T,-)
U+06FB
U+06FD
U+06FE
U+075C
U+0763
U+0768
U+076A
U+076B
U+076D
U+076E
U+076F
U+0770
U+0771
U+0772
U+0773
U+0774
U+0775
U+0776
U+0777
U+0778
U+0779
U+077A
U+077B
U+077C
U+077D
#

# Characters from Unicode Basic Latin Table (0000-007F):
U+002D
U+0061
U+0062
U+0063
U+0064
U+0065
U+0066
U+0067
U+0068
U+0069
U+006A
U+006B
```

```
U+006C
U+006D
U+006E
U+006F
U+0070
U+0071
U+0072
U+0073
U+0074
U+0075
U+0076
U+0077
U+0078
U+0079
U+007A
U+0030;U+06F0 (T,-)
U+0031;U+06F1 (T,-)
U+0032;U+06F2 (T,-)
U+0033;U+06F3 (T,-)
U+0034;U+06F4 (T,-)
U+0035;U+06F5 (T,-)
U+0036;U+06F6 (T,-)
U+0037;U+06F7 (T,-)
U+0038;U+06F8 (T,-)
U+0039;U+06F9 (T,-)
#

# Characters from Unicode General Punctuation Table (2000-206F):
U+200C
#

# Characters from Arabic language table
U+0629;U+06C3     Arabic Letter TEH MARBUTA
U+0643;U+06A9;U+06AA    Arabic Letter KAF
U+0649;U+064A;U+06CC;U+06CD   Arabic Letter ALEF MASKURA
U+0660;U+06F0 (T,-) Arabic-Indic Digit Zero
U+0661;U+06F1 (T,-) Arabic-Indic Digit One
U+0662;U+06F2 (T,-) Arabic-Indic Digit Two
U+0663;U+06F3 (T,-) Arabic-Indic Digit Three
U+0664;U+06F4 (T,-) Arabic-Indic Digit Four
U+0665;U+06F5 (T,-) Arabic-Indic Digit Five
U+0666;U+06F6 (T,-) Arabic-Indic Digit Six
U+0667;U+06F7 (T,-) Arabic-Indic Digit Seven
U+0668;U+06F8 (T,-) Arabic-Indic Digit Eight
U+0669;U+06F9 (T,-) Arabic-Indic Digit Nine
#
```

Annex-G: Character Sets of Pakistani Languages Provided by Language Representatives in the Meeting

Balochi

Balochi

ا ب پ ت ٹ
ج چ ح خ د ڈ
ر ز س ش ں ص ض
ع غ ف ک گ ل م ن
و ہ ء ی ے

ضیاء الحسن بلوچ
AIOU

Brahvi

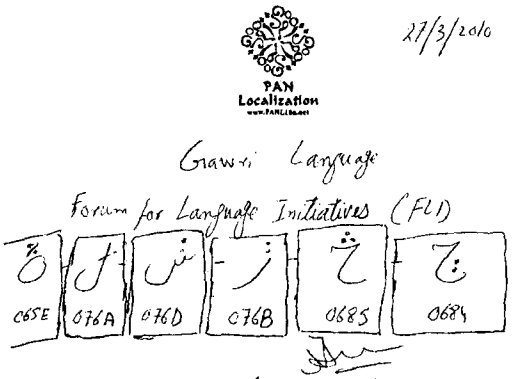
Brahvi

ا ب پ ت ٹ ج چ ح خ
د ڈ ر ڑ ز س ش ں ص ض
ط ظ ع غ ف ک گ
ل ڵ م ن و ہ ء ی

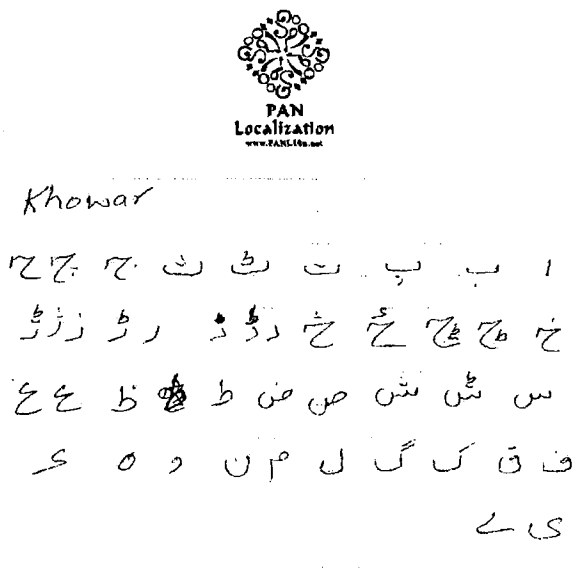
براہوی زبان میں ہائے آواز میں بھی ہیں جیسے
بھ پھ تھ ٹھ جھ چھ

ضیاء الحسن بلوچ
AIOU

Gawri



Khowar



Pashto

PAN Localization

پشتو الف با

ا ب پ ت ټ ث ج ځ چ څ

ح خ د ډ ذ ر ړ ز ږ س ش ښ

ص ض ط ظ ع غ ف ق ک ګ

ل م ن ڼ و ه ۀ ی ې ي ے

ئ

عبدالله جان عابد

لیکچرر

ڈیپارٹمنٹ آف پاکستانی لینگویجز

علامہ اقبال اوپن یونیورسٹی، اسلام آباد

Saraiki

PAN Localization

Saraiki Language

ا ب پ ت ٹ ث ج ڄ

چ ح خ د ڈ ذ ر ڑ ز ژ س ش ص ض

ط ظ ع غ ف ق ک گ گ

ل م ن ں ڻ و ھ ۂ ء ی ے ۓ

ڈاکٹر علمدار حسین بخاری

Dr. Alamdar Hussain Bukhari

Director

Saraiki Area Study Centre

(SASC) Bahaudin Zakariya

University Multan

Sindhi

1/2

PAN Localization

سندھی الفابیٹ کے مکمل حروف کا چارٹ

ا (آ الف مد)

ب ٻ ڀ پ ت ٿ ٽ ث ٺ

ف ڦ

ک ڪ گ ڳ ڱ

ج ڄ جھ ڃ چ ڇ ح خ

د ڌ ڏ ڊ ڍ

ر ڙ ز

ق ڪ

س ش ص ض ط ظ ع غ

ل م ن ڻ

و ھ ء ي

سفارشات

Missing letters in common language Table of Pakistani Language, are:

① (ھ) code 0647 (Already in Unicode table is existed)

② (جھ), (گھ) (These are complete characters of Sindhi language, not "وسرگ" (aspirade) sounds of 'ج' and 'گ' like 'Urdu')

So these letters may be included as

2/2

PAN Localization

seperate letters "جھ" & "گھ" in common language table or Key board.

④ Sindhi comma is like English comma, but its shape in Sindhi is "،". This shape of comma may be included in Key-board & common language table.

⑤ (ھ) last shape of 'ھ' like [illegible] may be adjusted in key board or language Table

(Taj Joyo)
Secretary Sindhi Language Authority
Hyderabad, Sindh