Urdu Domain Names

Sarmad Hussain and Nadir Durrani
Center for Research in Urdu Language Processing, National University of Computer and Emerging Sciences
B Block, Faisal Town, Lahore, Pakistan
{sarmad.hussain, nadir.durrani}@nu.edu.pk
ووو.اردوتحقیق.نیٹ or www.crulp.org

Abstract –With of international standards, including Unicode, CLDR, HTML, etc., it is now becoming increasingly possible to develop and deploy online content in local languages across the globe. However, a user is still required to write the domain name in Latin script to access this information on the internet, which still a barrier for non-Latin script based language speakers. This paper overviews the emerging Internationalized Domain Name (IDN) standards being proposed by Internet Corporation for Assigned Names and Numbers (ICANN). The paper also discusses challenges for implementing IDN for Urdu and a possible solution which has been implemented and is currently deployed.

Keywords: Urdu, IDN, Urdu Normalization, Urdu gTLD, Urdu ccTLD

I. Introduction

Language still remains one of the most formidable barriers to access of information through the internet. With advent of international standards, including Unicode [1], Common Locale Data Repository (CLDR) [2], HTML, etc., it is now becoming increasingly possible to develop and deploy content in local languages across the globe. This is providing access to populations which do not understand English or other foreign languages. However, even though it is possible to develop web pages in local languages, it is still not possible to easily access them without knowing Latin script and English conventions because the Domain Name System (DNS) is in Latin script and uses English-style conventions and abbreviations. One of the main reasons for this bottleneck is that the current Internet Protocol (IP) maps onto an addressing system that is based on the 7-bit ASCII standard and, therefore, it is not possible to encode multiple languages which would require the 16-bit Unicode standard[1]. There are two possible solutions to address this bottleneck: (i) develop systems which work independently of the existing DNS and, (ii) develop systems which work within the existing DNS. This has significant political, social and economic consequences, as currently the private consortium controlling the internet, ICANN, is based in US. See [3] for a more comprehensive overview.

II. ICANN's IDN in Applications (IDNA)

As discussed, the original DNS protocol was initially designed for ASCII character set. The relevant function gethostbyname() only allows ASCII. ICANN has been working on developing a system for IDN.

This solution adds a layer between DNS and the client at the application side, known as IDN in Application (IDNA) [4]. This layer takes the domain name in local language, normalizes it through *nameprep* process [5], and converts this non-ASCII string to a DNS compatible *ASCII Compatible Encoding* (ACE) known as *Punycode* [6]. This ensures backward compatibility. The DNS protocol continues to resolve the ASCII based domain name and get the IP address of host. This is illustrated in Figure 1 below.

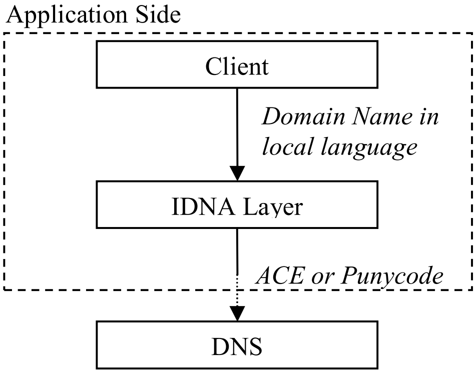

Fig. 1. Schematic for Conversion from IDN to ACE

IDNA layer is to be embedded within client side applications, e.g. the web browser, proxy server, etc. Details of this two step conversion are given below.

A. Nameprep Function

This function takes a string in local language and converts it into a normalized Unicode string. The string in local language may be using different encoding schemes e.g. UTF-8, ISO 8859-x, Unicode, Big5 (for Chinese), etc. The first step is to recognize the encoding and convert it into Unicode standard encoding, if required.

Unicode standard has redundancy within the standard, built in for backward compatibility and other reasons. Thus, the Unicode string has to be normalized in the second step of the process. For example, á (U+00E1) can also be written as a combination of a and ´ (U+0061 + U+0301). Details of Unicode normalization are given in [7]. For certain scripts, other considerations may also need to be taken. *Nameprep* is based on *stringprep* algorithm for internationalized strings [8].

[1] Even with Unicode there would be issues, as it is a script based standard.

B. *Punycode*

In order to make the hostname DNS compatible the Unicode string has to be converted to ACE. Many schemes have been proposed in this regard. Punycode is a boot-string encoding mechanism that uniquely converts Unicode string to the allowed ASCII based encoding. This conversion takes place through an algorithm known as ToASCII(). ToUnicode() converts back ASCII based encoding into Unicode compatible scheme. Punycode uses conventional ASCII i.e. a-z, 0-9 and hyphen, for backward compatibility [6].

The ToASCII function is applied separately to all the labels in domain name. There is a possibility that the generated Punycode is already a registered domain. For example, when http://www.ا.com is converted using ToASCII() function, the domain name ا (U+0627) is converted to ASCII string "mgb" but http://www.mgb.com may already be a registered domain. To avoid significant duplication, all such conversions through the ToASCII() function are appended with a four character prefix "xn--". The URL http://www.ا.com is therefore converted into http://www.xn--mgb.com.

III. Evaluation of DNS and IDNA

Apart from political issues, there are also some additional criticism associated with DNS system generally and specifically for IDNA.

Limitation of DNS to encode many languages due to its ASCII base has already been discussed.

RFC 920 [9] expanded the addressing convention to include top level domains (TLDs) like .edu, .com, .org, etc. However, now these are being used beyond the intended usage. For example, .com.la was sold by Lao PDR to a group which is using this TLD for Los Angeles city, and Tuvalu's county code .tv is being used by names associated with television.

It is currently being debated whether IDN should ride over the existing DNS system, as discussed. ICANN argues for the importance of a single root. However, there are also other parallel namespaces which are successfully working, e.g. for companies like AOL and Skype, and for countries like China Internet Network Information Center (CNNIC), Japan Network Information Center (JPNIC) and Korea Network Information Center (KRNIC), which are maintaining thousands of addresses. And other domains like telephone exchanges have shown that parallel systems can co-exist and effectively communicate with proper collaboration [3].

Though a unique Internet Protocol (IP) number identifies each address, DNS was introduced for the ease of users. However, DNS is not always visually unique, which can cause malicious or unintentional intervention. For example, lower case "L" in English looks similar to the upper case "I" or the digit "1" in some fonts. Thus, the website www.paypal.com may be written in different ways which are visually identical. Same is the case with the digit "0" and the upper case letter "O". This confusion could be even more profound if additional scripts are incorporated in the URL to enable multilingual IDN and more *phishing* [10] attacks are possible [3].

A variety of solutions have been considered to control the confusion that is caused by the visual similarity within a script, and enhanced by allowing multilingual domain names. At least one way to restrict some confusion is to disallow use of characters from different script blocks in Unicode to be used within a domain name. Thus, purely Arabic script domain names may be allowed but domain names with Arabic letters mixed with Latin characters may not be allowed. However, some languages do traditionally use some letters across scripts (as encoded in Unicode) and therefore some mechanism still needs to allow non-arbitrary and pre-defined mix of characters for certain languages. For example, Urdu may use digits in Latin block.

Moreover, there may be confusion for a language within a script block. Unicode being a script based standard groups all letters across all languages which use the same script. There also additional variants due to other reasons[2]. Thus, from a single language there may be redundancy. So beyond normalization [normalization], which is not language specific, further language dependent mapping may also be required.

Thus, language specific conventions need to be given for controlling which characters may be allowed within and across scripts for a particular language. This may also depend on where the language is used (for example, same language may be written using a different script in different regions). So the language specific information also needs to specify the region for which the conventions are valid. This may be achieved through defining language tables [17]. There tables are to be maintained by the registrars of domains. The table for each language would list the "base characters" it allows and their "variant(s)." In addition, it would also contain letters from other scripts conventionally used by the language. The language table is labeled with language and regional codes, e.g. those used in locale definitions [2]. See [16] for a template for defining a language table.

Finally, even though Punycode gives a unique mechanism for conversion between ASCII and Unicode, it is still being debated if this conversion will only be applicable the unique address or also to gTLDs and ccTLDs. This has significant political implications as well.

IV. Urdu Domain Names

Enabling domain names in Urdu also has significant political, social and financial implications. The rest of this paper discusses the technical challenges related to enabling Urdu IDN and proposes a solution. This solution has also been implemented as a concept system for testing and further improvement.

[2] For example, for backward compatibility.

A. Character Set

Urdu character set has been defined and standardized at national level [12, 13] and within Unicode [14]. In the character set there are different types of characters. These include basic alphabet, digits, vowel marks, punctuation marks, honorifics, and special symbols. See [13] for details. The first decision which needs to be made is which subset is allowed to be part of Urdu domain name. Latin based URLs allow "LDH" scheme, allowing letters 'a-z', digits '0-9' and hyphen '-'. Urdu has more complex writing conventions. At least, all the basic characters and digits must be allowed. Urdu also optionally uses diacritics, which help in defining the vowels. In normal writing these vowels are not written. However, they are used to disambiguate homographs (which are spoken in multiple ways but only distinguished based on the diacritics used). Also, certain diacritics are not optional and must be used for correct spelling, e.g. Khari Zabar (e.g. اعلیٰ) and Do-Zabar (e.g. تقریباً). If diacritics are allowed in the URLs, it would not be clear what would be the URL if the optional diacritics are not used. Urdu speakers would generally consider URLs with and without the optional diacritics equivalent. Thus, these optional diacritics are not required. However, non-optional diacritics would be expected by the Urdu readers, e.g. for the words given earlier. Study of a 12 million word corpus[3] of Urdu show that about 710 words with these required diacritics have occurred a total of 24,293 times. Generally, in this corpus optional diacritics were used 41,332 times showing a very small percent of words are typed with diacritics. This data shows that the use of diacritics may not be necessary. Additionally, the data shows that for the required diacritics, significant times the diacritic(s) are not placed consistently at the same place, e.g. الہیٰ , الہٰی and الٰہی (found 11, 50 and 549 times in the corpus respectively). Thus, if they are allowed, it may introduce another way of phishing.

Most of the pronunciation marks are not necessary for URL and may be excluded. However, Urdu end of sentence marker '۔' is needed to separate the domain name, gTLD and ccTLD. This has two associated issues. First, should it be synonymously used with the Latin period '.'? Second issue is that this end of sentence marker for Urdu is a homograph of hyphen, which is allowed in URLs but does not act as a separator between domain names, gTLDs and ccTLDs. Thus, it would become very confusing for the user when period, hypen and end of sentence marker for Urdu are mixed, but would be entirely possible for multilingual domain names. An added problem may occur when the period is mixed with Urdu digit zero, which is almost a homograph. ".-٠۔" shows a Period-Hypen-UrduZero-UrduEndOfSentenceMarker sequence. A solution is not to allow hyphen in Urdu domain names and allow Urdu end-of sentence marker to be used synonymously with '.' as tag separators within a domain name.

Honorifics are optional in most cases, or have a regular (longer character based phrasal equivalent). For Muslims, it is recommended to put "ؐ" symbol or the equivalent "£" ligature[4] with the name of Prophet Muhammad. If it is disallowed in the domain name, then any website which uses this address would need to have the fully expanded form, which may be very long and difficult for users to type out. Two other honorifics, 'ؓ' and 'ؑ', are also mandatory when mentioning names of the companions of Prophet Muhammad and other prophets respectively. Other honorifics are optional. These honorifics have been used 292 times in the 12 million word Urdu corpus. It is recommended to allow required honorifics as the variant forms, which can be de-normalized as given in Table 3 in Appendix B. Others optional honorifics e.g. 'ؒ' and 'ﷻ'may be disallowed.

Other symbols are mostly notational (e.g. footnote marker, sign to indicate a verse quoted in prose, etc.). These are not necessary for inclusion in the domain names.

B. Cursiveness

Arabic writing system is highly cursive, with most letters having at least four shapes, when they occur in the beginning, middle, and end of a sequence and in isolation. There are two kinds of letters, one set which can join with others, and another set of letters which cannot join with letters after them. As domain names do not allow the space character within them, if multiple words of Urdu are written, they would join together and may be mis-read. In English, words can be separated by hypen or using a capital letter, e.g. "two-words" and "TwoWords" so space is not required. However, Urdu neither has a hyphen nor capital letters. It is possible to insert a zero-width-non-joiner (ZWNJ, U+200C) but this character is not familiar for users. The second option is to allow for the space character by the users for proper visual rendering of multiple words. The space may be removed in the *nameprep* or other process at client side or at the registrar, so that the final Unicode output would not include it. Similar treatment may be done with ZWNJ. However, this will allow the user to view the domain name correctly.

C. Encoding

Minimally Urdu Zabta Takhti [13], UTF-8 and Unicode support must be provided. There are also other non-standard encodings but should remain out of scope of this process.

D. Normalization

There is a lot of redundancy in Unicode for Arabic script. Arabic block is from U+0600 till U+06FF and

[3] This unpublished corpus is balanced over different genres and is derived from online material published after 1996.

[4] This symbol stands for the phrase "peace be upon him." A third variation is to actually spell out the whole phrase.

extended Arabic from U+0750 to U+077F. In addition, for backward compatibility, actual position based glyphs have also been included from U+FB50 till U+FDC7 and U+FE70 till U+FEFF. Finally, special symbols are listed at U+FDFx.

Three kinds of normalization are required. First, there are characters within the Unicode which are repeated for different languages allowing redundancy. This redundancy must be removed to allow unique naming space. For example, there are two sets of digits, one for Arabic and other for remaining languages (e.g. Farsi, Urdu, Sindhi, etc.). However, though the following are written using these two different sets of Unicode values, ١٢٣ and ۱۲۳ are visually same. A complete list of potentially confusing characters from the perspective of Urdu and their recommended equivalents for normalization are given in Appendix B. In addition, all the Arabic Presentation Forms should also be mapped onto the base forms within U+06xx. However, not all characters are easily possible to map. Some characters do not share the same behavior but are still confusing and may be used naïvely or maliciously. Thus the normalization process needs to be extended beyond the permitted canonical limits proposed by Unicode to prevent these possibilities. Thus, all types of *Yay*, *Hay*, etc. are normalized. It is also important to note that these normalizations would not work across other languages (e.g. Sindhi, Pashto, etc.) and are only done in context of Urdu. Thus, these must be included in the language table at the registrar, as proposed by [17]. The “base characters” are given in Appendix A and the one’s that have variants are listed in Table 1 in Appendix B.

Second, when base letters combine with some combining characters, their equivalent is also encoded directly in Unicode. Thus, آ can be written as U+0622 or a combination of U+0627 and U+0653. However, these sequences should be normalized. This normalization is also given in Table 2 in Appendix B and is part of the *nameprep* process.

Finally, Unicode also lists many ligatures. These ligatures must be de-normalized into base characters as well. A list of ligatures and their character equivalents is also given in Table 3 in Appendix B. These must also be done in the *nameprep* procedure.

E. Writing Style

Though Naskh style of writing is acceptable, Urdu language speakers prefer Nastalique style of writing [hussain3]. This is a font issue and though it has implications on the client side graphical user interface, it has no implication on IDN or *nameprep* function.

F. gTLDs

Urdu would eventually need its own gTLD set and separate name space. However, along with that it should also be possible to access existing namespaces in Urdu using direct mapping. Latter is already possible, if incorporated at the client side during *nameprep* processing. A set of gTLDs and their translation are listed in Appendix C which may be used for such mapping.

G. ccTLDs

Similar to gTLDs, ccTLDs also need to be translated. However, they would share the same namespace and must be mapped onto existing ccTLDs at the client side. However, Urdu translation is still required for the Urdu users. The mapping is given in Appendix D.

G. Conversion of www

As for the regular URLs, Urdu address would also need to specify the name space. The first portion of the string normally specifies www. This could be transliterated into Urdu as ووو to represent the same space. However, it would need to be transliterated to www at the client side for further processing.

V. Urdu Domain Name Algorithm

A solution for Urdu domain names would have the following steps:

1. Use the separator and divide the URL into different portions
2. Convert ووو to www
3. Remove Diacritics, honorifics and any special symbols from the domain address, except the honorific used for Prophet Muhammad, his companions and earlier prohpets
4. Remove space or ZWNJ markers from the domain address
5. Normalize the resulting domain address using the rules discussed above
6. Use ToASCII() function to generate the Punycode equivalent string
7. Check if the Punycode is from valid characters using the language tables at the registry, after variant characters are mapped onto the base characters
8. Find English mapping of the Urdu gTLD
9. Find English mapping of the Urdu ccTLD
10. Keep any trailing string unchanged
11. Concatenate the strings from 2, 6, 7, 8 and 9 to form the corresponding English URL
12. Forward the address as an http request

Steps 8 and 9 is currently doing simple mapping from Urdu to English equivalents. However, if local language gTLDs are also enabled, then Punycode conversion would be required at this step instead of mapping.

This procedure converts the Urdu domain name ووو-اردوتحقیق-نیٹ into www.xn--mgbgjgj9ha8b83g.net[5], and

[5] The conversion has been done from the site http://www.panl10n.net/Puny/udnc.php. Also see

does not allow spurious domain names, to avoid confusion. The procedure is as per the ICANN guidelines [18].

VI. Issues and Future Considerations

Spoofing [11] and Phishing [10] attacks are one of the major concerns for IDNA. As discussed, people can use homoglyphs (or visually similar characters) to develop alternate websites which look similar to target websites. These spoofed websites can then phish for users to acquire private information of the users (e.g. credit card information etc.). Like other languages, Urdu IDNA system will also be open to such attacks. The extended normalization process suggested above has been devised to rebuff such possibilities. However, there are still other ways to spoof using other language characters. Thus such possibility cannot be totally controlled. Other mechanisms need to be developed for better control, e.g. security certificates, filters, etc. A complete discussion is beyond the scope of the paper. However, see [10, 11] for further details.

Even after normalization, it is also possible to "spoof" using legitimate means due to spelling variations and other methods. Thus, *www.color.com*, *www.colour.com*, *www.color.net* are all possible spoofing possibilities for *www.colour.net*. Similar possibilities also exist in Urdu. For example, ووو.زکوت.نیٹ may be confused with ووو.زکوۃ.نیٹ even though the two are legitimately different based on encoding and confusion occurs based on how Urdu uses these characters.

However, technical and security constraints cannot undermine the immense potential and requirement of localized domain names. It is necessary for bridging the digital divide and give access to the universe of online content to local populations. Both public and private organizations need to strive to provide a secure but universal access to cyberspace [unesco].

Acknowledgements

This work carried out with the aid of a grant from International Development Research Center (IDRC), Ottawa, Canada, administered through the Center for Research in Urdu Language Processing (CRULP), National University of Computer and Emerging Sciences (NUCES), Pakistan. The authors would also like to acknowledge the support of researchers at CRULP for providing data and valuable feedback on the paper.

References

[1] Unicode Consortium, *Unicode 5.0,* 5th ed., Addison-Wesley Professional, 2006.

[2] Unicode Consortium, "Common Locale Data Repository Project," http://www.unicode.org/press/pr-cldr1.4.html

[3] D. Butt, "Internationalized Domain Names," http://www.apdip.net /apdipenote/9.pdf, APDIP.net, 2006.

[4] P. Faltstrom, P. Hoffman, and A. Costello, "Internationalizing Domain Names in Applications (IDNA)," http://www.rfc-editor.org/rfc/rfc3490.txt, 2003.

[5] P. Hoffman,M. B. Viagenie, "Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN)" http://www.rfc-editor.org/rfc/rfc3491.txt, 2003.

[6] A. Costello, "Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)," http://www.rfc-editor.org/rfc /rfc3492.txt, 2003.

[7] M. Davis, M. Durst, "Unicode Normalization Forms," http://www.unicode.org/reports/tr15/, 2005.

[8] P. Hoffman and M. B. Viagenie, "Preparation of Internationalized Strings ("stringprep")," http://tools.ietf.org /html/rfc3454, 2002.

[9] J. Postel, J. Reynolds, "Domain Requirements," http://www.rfc-editor.org/rfc/rfc920.txt, 1984.

[10] Wikipedia, "Phishing," http://en.wikipedia.org/wiki /Phishing.

[11] Wikipedia, "Spoofing Attack," http://en.wikipedia.org /wiki/ Spoofing_attack.

[12] M. Afzal and S. Hussain, "Urdu Computing Standards: Development of UZT 1.01," in *Proceedings of the IEEE International Multi-Topic Conference*, Lahore, 2001.

[13] . S. Hussain and M. Afzal, "Urdu Computing Standards: UZT 1.01," in *Proceedings of the IEEE International Multi-Topic Conference,* Lahore, 2001.

[14] S. Hussain & K. Zia, "Proposal to Add Marks and Digits in Arabic Code Block (for Urdu)," in *Proceeding of 42nd Meeting of ISO/IEC JTC1/SC2/WG2,* Dublin, Ireland, 2002.

[15] S. Hussain, "Complexity of Asian Scripts: A Case Study of Nafees Nasta'leeq," in the *Proceedings of SCALLA*, Kathmandu, Nepal, 2004.

[16] http://www.iana.org/assignments/idn/registry-language -template.txt

[17] J. Klensin, "Suggested Practices for Registration of Internationalized Domain Names (IDN)," http://www.rfc-editor.org/rfc/rfc4290.txt, 2005.

[18] "Guidelines for Implementation of Internationalized Domain Names," http://www.icann.org/general/idn-guidelines-22feb06.htm, 2006.

Appendix A: Base Character Set

Glyph	Unicode	Glyph	Unicode

http://mct.verisign-grs.com/index.shtml for similar purpose, though latter is not completely functional for Urdu.

۰	06F0	ش	0634
۱	06F1	ص	0635
۲	06F2	ض	0636
۳	06F3	ط	0637
۴	06F4	ظ	0638
۵	06F5	ع	0639
۶	06F6	غ	063A
۷	06F7	ف	0641
۸	06F8	ق	0642
۹	06F9	ک	06A9
ا	0627	گ	06AF
آ	0622	ل	0644
ب	0628	م	0645
پ	067E	ن	0646
ت	062A	ں	06BA
ٹ	0679	و	0648
ث	062B	ہ	06C1
ج	062C	ھ	06BE
چ	0686	ۃ	06C3
ح	062D	ء	0621
خ	062E	ی	06CC
د	062F	ے	06D2
ڈ	0688	أ	0623
ذ	0630	ؤ	0624
ر	0631	ۓ	06D4
ڑ	0691	ئ	0626
ز	0632	ۓ	06D3
س	0633	ۂ	06C2

APPENDIX B: LETTER NORMALIZATION FOR URDU

Table 1: Letter normalization for Urdu

Variant Form	Recommended Base Form
١ (661)	۱(6f1)
٢(662)	۲(6f2)
٣(663)	۳(6f3)
٤(664)	۴(6f4)
٥(665)	۵(6f5)
٦(666)	۶(6f6)
٧(667)	۷(6f7)
٨(668)	۸(6f8)
٩(669)	۹(6f9)
٠(660)	۰(6f0)
ك (643)	ک (6A9)
ى(649)	ی(6CC)
ى(649)	ي(64A)
ة (629)	ۃ (6C3)

Table 2: Case fold normalization for Urdu

Characters	Normalized Form	Recommended Form
ٓ+ا	آ	آ
ء+ا	أ	أ
ء+و	ؤ	ؤ
ء+ے	ۓ	ۓ
ء+ہ	ۂ	ۂ
ء+ی	ئ	ئ

Table 3: Ligature normalization for Urdu

Ligature Form	Recommended Form
لا	ا+ل
الله	ه+ل+ ل + ا
□	د+م+ح+م
اکبر	ر+ب+ک+ا
صلعم	م+ع+ل+ص
رسول	ل+و+س+ر
علیہ	ه+ی+ل+ع
وسلم	م+ل+س+و
صلی	ی+ل+ص
□	+م+ ح + ا ر+ه+ ل + ل+ ا+ م س+ب +ی م+ ح + ر + ا+ن
£	+س+و+ه+ی+ل+ع+ه+ل+ل+ا+ی+ل+ص م+ل

ﷺ	ص+ل+ی+ا+ل+ل+ہ+ع+ل+ی+ہ+و+س+ل+م
ﷻ	ج+ل+ج+ل+ا+ل+ہ

APPENDIX C: URDU TRANSLATIONS FOR GTLD

English gTLD	Abbr. of gTLD	Urdu gTLD
APRA	apra	انٹرنیٹ
COMPANY	com	کمپنی
EDUCATION	edu	تعلیم
GOVERNMENT	gov	حکومت
MILITARY	mil	فوج
ORGANIZATION	org	ادارہ
INTERNATIONAL	int	عالمی
NET	net	نیٹ
INFORMATION	info	اطلاعات
MEDIA	media	میڈیا
NAME	name	نام
BUSINESS	biz	کاروبار
AEROSPACE	aero	فضائیات
PROFESSIONAL	pro	پروفیشنل
COOPERATIVE	coop	کواپریٹو
MUSEUM	museum	میوزیم
Employment Related	jobs	ملازمت
Travel Agents, Airlines	travel	سیاحت
Asian Community	asia	ایشیا
Mobile Devices	mob	موبائل
Telephone network and Internet	tel	مواصلات
Postal Service	post	ڈاک
Government of Sindh	GOS. PK	سندھ حکومت
Government of Punjab	GOP. PK	پنجاب حکومت
Government of NWFP	GON. PK	سرحد حکومت
Government of Azad Kashmir	GOK. PK	کشمیر حکومت
Government of Baluchistan	GOB. PK	بلوچستان حکومت
Web Sites	Web	ویب
Family and Individuals	fam	خاندان

APPENDIX D: TRANSLATIONS FOR CCTLD'S

English ccTLD	Abbr. for cc-TLD	Urdu ccTLD
Afghanistan	af	افغانستان
Albania	al	البانیہ
Algeria	dz	الجیریا
American Samoa	as	امریکی سموآ
Andorra	ad	انڈورا
Angola	ao	انگولا
Anguilla	ai	انگوایلا
Antarctica	aq	انٹارکٹکا
Antigua and Barbuda	ag	اینٹی گوا اینڈ باربوڈا
Argentina	ar	ارجنٹائن
Armenia	am	آرمینیا
Aruba	aw	اروبا
Australia	au	آسٹریلیا
Austria	at	آسٹریا
Azerbaijan	az	آزربائیجان
Bahamas	bs	بہا ماس
Bahrain	bh	بحرین
Bangladesh	bd	بنگلہ دیش
Barbados	bb	باربادوس
Belarus	by	بیلارس
Belgium	be	بیلجیم
Belize	bz	بیلز
Benin	bj	بینن
Bermuda	bm	بینجمن
Bhutan	bt	بھوٹان
Bolivia	bo	بولیویا
Bosnia and	ba	بوسنیا

Herzegovina		
Botswana	bw	بوٹسوانا
Bouvet Island	bv	بوئیٹ آئ لینڈ
Brazil	br	برازیل
Brunei	bn	برونائی
Bulgaria	bg	بلغاریہ
Burkina Faso	bf	برکینا فاسکو
Burundi	bi	برونڈی
Cambodia	kh	کمبوڈیا
Cameroon	cm	کیمرون
Canada	ca	کینیڈا
Cape Verde	cv	کیپ ورڈے
Cayman Islands	ky	کیمین جزائر
Central African Republic	cf	وسطی افریقہ
Chad	td	چاڈ
Chile	cl	چلی
China	cn	چین
Christmas Island	cx	کرسمس آئی لینڈ
Cocos (Keeling) Islands	cc	کوکوز جزائر
Colombia	co	کولمبیا
Comoros	km	کوموروس
Congo	cd	کانگو
Cook Islands	ck	کک آئی لینڈز
Costa Rica	cr	کوسٹاریکا
Cote d'Ivoire	ci	آئیوری کوسٹ
Croatia	hr	کروشیا
Cuba	cu	کیوبا
		قبرص

Cyprus	cy	
Czech Republic	cz	جمہوریہ چیک
Denmark	dk	ڈینمارک
Djibouti	dj	جبوتی
Dominica	do	ڈومینیکا
East Timor	tp	مشرقی تیمور
Ecuador	ec	ایکواڈور
Egypt	eg	مصر
El Salvador	sv	ایسلواڈور
Equatorial Guinea	gq	گنی استوائی
Eritrea	er	ایریٹریا
Estonia	ee	آسٹونیا
Ethiopia	et	ایتھوپیا
Falkland Islands	fk	فاک لینڈ جزائر
Faroe Islands	fo	فیرو جزائر
Fiji	fj	فجی
Finland	fi	فِن لینڈ
France	fr	فرانس
French Guiana	gf	فرنچ گوئیانا
French Polynesia	pf	فرنچ پولینسیا
French Southern and Antarctic Lands	tf	فرانسیسی قطبِ جنوبی
Gabon	ga	گبیون
Gambia	gm	گیمبیا
Gaza Strip		گولی
Georgia	ge	جارجیا
Germany	de	جرمنی

Ghana	gh	گھانا
Gibraltar	gi	جبرالٹر
Greece	gr	یونان
Greenland	gl	گرین لینڈ
Grenada	gd	گریناڈا
Guadeloupe	gp	گواڈی لوپ
Guam	gu	گوام
Guatemala	gt	گوئٹے مالا
Guernsey	gg	گونیسی
Guinea	gn	گنی
Guinea-Bissau	gw	گنی بساؤ
Guyana	gy	گوآنا
Haiti	ht	ہیٹی
Heard Island and McDonald Islands	hm	ہیملٹن
Holy See (Vatican City)	va	ویٹیکن سٹی
Honduras	hn	ہنڈوراس
Hong Kong	hk	ہانگ کانگ
Hungary	hu	ہنگری
Iceland	is	آئس لینڈ
India	in	انڈیا
Indonesia	id	انڈونیشیا
Iran	ir	ایران
Iraq	iq	عراق
Ireland	ie	آئر لینڈ
Israel	il	اسرائیل

Italy	it	اٹلی
Jamaica	jm	جمیکا
Jan Mayen	sj	جان مائین
Japan	jp	جاپان
Jersey	je	جرسی
Johnston Atoll		جانسٹن آٹول
Jordan	jo	اردن
Kazakhstan	Kz	قازقستان
Kenya	ke	کینیا
Kiribati	ki	کریباتی
North Korea	kp	شمالی کوریا
South Korea	kr	جنوبی کوریا
Kuwait	kw	کویت
Kyrgyzstan	kg	کرغستان
Laos	la	لاؤس
Latvia	lv	لٹویا
Lebanon	lb	لبنان
Lesotho	ls	لیسوتھو
Liberia	lr	لائبیریا
Libya	ly	لیبیا
Liechtenstein	li	لشٹنسٹائن
Lithuania	lt	لیتھوآینا
Luxembourg	lu	لیکسمبرگ
Macau	macau	مکاؤ
Macedonia	mk	مَیک ڈونلڈ
Madagascar	mg	مقدونیا
Malawi	mw	ملاوی
		ملائشیا

Malaysia	my	
Maldives	mv	مالدیپ
Mali	ml	مالی
Malta	mt	مالٹا
Marshall Islands	mh	مارشل آئی لینڈ
Martinique	mq	مارٹینیق
Mauritania	mr	موریطانیہ
Mauritius	mu	موریشس
Mayotte	yt	مے اوٹی
Mexico	mx	میکسیکو
Micronesia	fm	مائیکرونیسیا
Moldova	md	مالڈوا
Monaco	mc	مناکو
Mongolia	mn	منگولیا
Montserrat	ms	ماونٹ سیرت
Morocco	ma	مراکش
Mozambique	mz	موزمبیق
Namibia	na	نمیبیا
Nauru	nr	ناؤرو
Nepal	np	نیپال
Netherlands	an	نِیدرلینڈز
New Caledonia	nc	نیو کیلی ڈونیا
New Zealand	nz	نیوزی لینڈ
Nicaragua	ni	نکاراگوا
Niger	ne	نائجیر
Nigeria	ng	نائجیریا
Niue	nu	نییو
Norfolk Island	nf	نورفوک آئی لینڈ

Northern Mariana Islands	mp	شمالی مرینا آئی لینڈز
Norway	no	ناروے
Oman	om	اومان
Pakistan	pk	پاکستان
Palau	pw	پلاؤ جزائر
Panama	pa	پانامہ
Papua New Guinea	pg	پاپوائے نیوگنی
Paraguay	py	پیراگوئے
Peru	pe	پیرو
Philippines	ph	فلپائن
Pitcairn Islands	pn	پیگیرن جزائر
Poland	pl	پو لینڈ
Portugal	pt	پرتگال
Puerto Rico	pr	پیورٹوریکو
Qatar	qa	قطر
Reunion	re	ری یونین
Romania	ro	رومانیہ
Russia	ru	روس
Rwanda	rw	روانڈا
Saint Helena	sh	سینٹ ہیلینا
Saint Kitts and Nevis	kn	سینٹ کیش اینڈ نیویز
Saint Lucia	lc	سینٹ لوئیسیا
Saint Pierre and Miquelon	pm	سینٹ پیری اینڈ میکولین
Saint Vincent and the Grenadines	vc	سینٹ ونسنٹ اینڈگریناڈائز
		ساموآ

Samoa	ws	
San Marino	sm	سان مرینو
Sao Tome and Principe	st	ساؤ ٹام اینڈ پرنشپ
Saudi Arabia	sa	سعودی عرب
Senegal	sn	سینیگال
Seychelles	sc	سیشیلز
Sierra Leone	sl	سیرالیون
Singapore	sg	سنگا پور
Slovakia	sk	سلواکیہ
Slovenia	si	سلوونیا
Solomon Islands	sb	سولومون جزائر
Somalia	so	صومالیہ
South Africa	za	جنوبی افریقہ
South Georgia and the South Sandwich Islands	gs	ساؤتھ جارجیا اینڈ ساؤتھ سینڈوچ آئی لینڈز
Southern Ocean		بحر منجمد جنوبی
Spain	es	سپین
Sri Lanka	lk	سری لنکا
Sudan	sd	سوڈان
Suriname	sr	سرینام
Swaziland	sz	سوازی لینڈ
Sweden	se	سویڈن
Switzerland	ch	سوئٹزرلینڈ
Syria	sy	شام
Tajikistan	tj	تاجکستان
Tanzania	tz	تنزانیہ
		تھائی لینڈ

Thailand	th	
Togo	tg	ٹوگو
Tokelau	tk	ٹوکیلاؤ
Tonga	to	ٹونگا
Trinidad and Tobago	tt	ٹرمینیدادا اینڈ ٹوباگو
Tunisia	tn	تیونس
Turkey	tr	ترکی
Turkmenistan	tm	ترکمانستان
Turks and Caicos Islands	tc	ترک اور کیکاؤس جزائر
Tuvalu	tv	ٹوالو
Uganda	ug	یوگنڈا
Ukraine	ua	یوکرائن
United Arab Emirates	ae	متحدہ عرب امارات
United Kingdom	uk gb	مملکت متحدہ برطانیہ
United States	us	امریکہ
Uruguay	uy	یوروگائے
Uzbekistan	uz	ازبکستان
Vanuatu	vu	وانوآتو
Venezuela	ve	وینزویلا
Vietnam	vn	ویت نام
Virgin Islands	vg	ورجن آئی لینڈز
Wallis and Futuna	wf	ویلس اینڈ فتونہ آئی لینڈز
Western Sahara	eh	مغربی صحارا
Yemen	ye	یمن
Zambia	zm	زمبیا
		زمبابوے

Zimbabwe	zw	
Taiwan	tw	تائیوان
European Union	eu	یورپین یونین
Myanmar	mm	میانمار
Palestinian State (proposed)	ps	فلسطینی ریاست (مجوزہ)