ICANN ICANN Email List Archives

L2/05-340

[idn-guidelines]

<<< Chronological Index >>> Thread Index >>>

domain name speculation

- *To*: <idn-guidelines@xxxxxxxxx>
- *Subject*: domain name speculation
- *From*: "Tim Hayward" <tim@xxxxxxxx>
- *Date*: Tue, 20 Sep 2005 23:21:55 -0500

It would be very cool if there was a way to decrease the attactivness of speculation in domain names. I was trying to set up a free, nothing but public service website today and I find that every two word name I think of has been taken for the sole purpose of making money on a supposed need. That is what has lead the way for urban sprawl in the US. Help fight it somehow.

Thks

Tim.

--

<<< Chronological Index >>> Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

Domain Forgery, Hijacking and False Whois

- *To*: idn-guidelines@xxxxxxxxx
- *Subject*: Domain Forgery, Hijacking and False Whois
- *From*: showker@xxxxxxxxxxx
- *Date*: Wed, 21 Sep 2005 13:04:52 -0400

```
This measure will tend to complicate the already
difficult task of tracking and apprehending online
criminals and criminal activities.

It will also further involve ICANN in complicity
in such online crimes by designating registrars
to allow serious violations of ICANN regulations,
and international law.

So far ICANN has NOT enforced its already existing
regulations and rules with registrars, and openly
and freely allows such registrars to engage in
numerous online criminal activities.

ICANN should be ELIMINATING loop-holes and the very
complexity which gives free reign to terrorists,
online criminals and get-rich profiteers.

ICANN should be working on setting up programs which
promote true authentication of domain users, and their
registrars -- along with providing more speedy remedy
for criminal activities online rather than making it
easier for the criminals to operate and more
difficult for law enforcement to track the
criminals down.

ICANN cannot even write a lucid draft proposal.

If ICANN cannot administrate the overwhelming glut
of bureaucracy it already has, what makes them think
they can handle adding more.

Regards
FNS, Spam Trackers
```

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

two quotes calling for more information

- *To*: idn-guidelines@xxxxxxxxx
- *Subject*: two quotes calling for more information
- *From*: r&d afrac <rd@xxxxxxxxx>
- *Date*: Fri, 23 Sep 2005 00:28:04 +0200

The reviewers write: "the IETF is currently dealing with the final draft of a successor document to that BCP. This will provide expanded means for specifying languages, including designations for script and orthographic authority as components of a language tag".

This is a very odd reading of the RFC 3066 bis Draft.

1) The "script" indication has strictly no interest as far as IDNs are concerned, except to provide a short (100 codes) list of names to possibly designate charsets (experts think there are around 200 scripts in use). For example, the debate on the Unicode list shown a deep disagreement over the simple question "can I write French in using the Latn script?".

This may introduce incredible confusion (we already met) as the same name can be used for different charsets by the TLD Managers. It is also of low interest because of the need to support foreign script characters used in commercial denominations, TM or nicknames. The need is for a naming format for TLD table.

2) I wander what can be called "orthographic authority" in the RFC 3066 language tag. That tag uses a mix of UN.49 and ISO 3166 codes and numbers to designate what they call a "region". The only "orthographic authority" could be the legal authority of the considered country. By nature the ccTLD (RFC 1591, GAC declaration) the ccTLD Manager as the trustee of the national authority IS that authority, its representative or its legal "colleague".

As a general comment, this WG decided non to consider IDNs, DNS, lingual community needs and convergence with other Registries (ISO 11179) as irrelevant to the scope of their WG. This will lead to appeals by lingual organisations and ccTLDs should this Draft be approved without supporting URI/IRI-tags. The conflict is between two different layers (computers interoperability and users interintelligibility):

- a sabilisation of the internationalisation approach in constraining the current pratice for a convergence (through the denomination) with locales files (this is the Unicode CLDR project quoted in the Charter)

- and a multilingualisation approach supporting the users lingual communities, including ccTLDs.

There are many reasons to use URI-tags (specified by an accepted non published RFC) and IRI-tags to be specified. One is homographs. The recommendations given by the Guide Lines only concerns SLDN. Phishing has nothing to do with IDNs but with the misconcepts of IDNA. Phishing uses IDNA possibilities on nLD labels and work with _every_ domain name. IRI-tags permit zone managers (any level) to specify their own tables for their own zone. A zone manager of the DN xxx.yyy.zz can define a Zone Table for Arab characters as "ar-0-xxx.yyy.zz:ar.txt".

The reviewers also note "The discussion that is in progress about permitting a more

extensive character repertoire in top-level labels can result in a change to this condition, as well as raising need for further guidelines specific to the new situation".

Could they document this "discussion in progress"?

JFC Morfin

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

Business and Intellectual Property Issues with the IDN Guidelines

- *To*: idn-guidelines@xxxxxxxxx
- *Subject*: Business and Intellectual Property Issues with the IDN Guidelines
- *From*: Danny Younger <dannyyounger@xxxxxxxxx>
- *Date*: Fri, 7 Oct 2005 06:47:16 -0700 (PDT)

I thank the ICANN Board for the opportunity to comment on the Draft Revised IDN Guidelines. I have one concern.

When policy is crafted, one must ensure that such policy does not unjustifiably injure particular categories of persons or entities. The Draft Revised IDN Guidelines as written have the potential to become policy which may significantly impact the worldwide business community (that seeks to promote its brands on an international basis) by restricting domain name registrations to a single language or a single script. This has the effect of denying a brand the possibility of launching a multilingual campaign on the Internet that can be coordinated with print and other media campaigns -- allow me to illustrate:

Let's suppose that NIKE (used only for the sake of an example) wants to run a worldwide promotion using the phrase "Run with Nike."
Their campaign would translate the words "run with" into a large number of languages, while the brand-word "Nike" remained as written in English. While they could display this message to the world in a multitude of languages on billboards, in newspapers and magazines, in TV and movie ads, they couldn't register any comparable domains in a multitude of multilingual forms because the single language/script rules won't allow them to do so.

I understand that the fear of homographic attacks has prompted the decision to restrict domains to a single language or script, yet I believe that this approach should be viewed as somewhat shortsighted. Trademarks all have one thing in common -- a recognition factor. A user can see a trademark symbol and instantly understand that the string of letters that preceded it is a trademark/brand.

I believe that an exception to the one language/script policy should be made when trademarks appear within domain names (as long as the trademark symbol is included after the name) -- unfortunately, the LDH code point approach currently excludes this possibility (as the TM symbol does not reside within this range).

I would argue that we should be able to technically provide for the needs of the intellectual property and business communities.

Providing this type of exception to the one language/script rule would allow registries and registrars to profit (as they could charge for the service of validating the trademarked string), the business community would benefit from the new opportunity that multilingual campaigns can offer, as would the users of such business products.

A validated trademark when combined with characters in another language/script does not pose a homographic threat (as long as the validation process confirms that the holder of the trademark is also the registrant of the domain).

The Internet is an economic engine. We shouldn't inadvertently craft policy that has the effect of being detrimental to economic growth.

Thank you for considering these remarks.

Best wishes,
Danny Younger

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

Comments on the Draft Revised IDN Guidelines

- *To*: idn-guidelines@xxxxxxxxx
- *Subject*: Comments on the Draft Revised IDN Guidelines
- *From*: Gervase Markham <gerv@xxxxxxxxxxx>
- *Date*: Tue, 11 Oct 2005 16:42:50 +0100

Dear ICANN,

I am writing to comment on the Draft Revised IDN Guidelines:
http://icann.org/general/idn-guidelines-20sep05.htm

Firstly, I would like to thank ICANN for taking a lead in tackling the potential spoofing problems that IDN entails, and in updating the guidelines. Overall, I am very pleased with the new draft, which represents a significant improvement over version 1.0.

From the point of view of the Mozilla Foundation, we would like the final draft document to be of sufficient clarity, extent and watertightness that we can simply say to registries "Do you follow all of the ICANN guidelines? If so, show us your documentation to prove it and we'll enable IDN for your TLD." I would hope that ICANN also has this aim - that is, that the guidelines should encapsulate the whole of current best practice in avoiding spoofing issues.

Let me focus my comments in by saying that I am in complete agreement with every guideline apart from guidelines 3) and 4). I think I am also in agreement with section 3), but there are some parts where I would seek clarification, and other parts where I think the guidelines do not go quite far enough. I will now elaborate on what I mean.

Guideline 3)

Reading Guideline 3 raises the following questions:

a) Who defines a "set of languages"? If the registries define them, what is to prevent a registry defining a set which contains every existing language, and therefore bypassing the intent of many of the guidelines? Is the number of necessary sets small enough to enumerate them in the document, as UTR #36 does[0]?

b) The following sentence is confusing to me:

"Visually confusable characters from different scripts may not appear in a single *label* unless there are overriding legitimate linguistic reasons for doing so."

Let's say that "b" and "6" are confusable, just for the sake of example. Taking this at face value, it says that I can't have a label like business60.com, because it has both a b and a 6 in it. This seems an odd thing to explicitly prohibit; surely the risk is not in having "business60.com", but in having both of "business60.com" and 6usinessb0.com" registered to different people?

Perhaps where *label* was written (highlighted above), you meant "table"? That would turn it into a very sensible restriction which made sense in the context.

c) As you will know, browsers do not have access to character tables or script labels

or any of this ancillary information about labels, and no-one (so far as I know) is proposing any mechanisms for them to have access to it. Therefore, any registry policy designed to prevent spoofing needs to be blind to the existence of character tables, even if they are used as a way of limiting registrations.

What I am saying here is that the current policy does not address whole-script spoofables (e.g. caxap.tld in Latin and caxap.tld in Cyrillic). If the .tld registry has a table for both Russian and English, nothing I can see in the guidelines tells them they need to make sure these two domains are not registered to different entities. In terms of the "safety" of these guidelines, this is my main concern.

d) Following on from the above, I believe it's important for transparency reasons for people (including browser manufacturers) to be able easily to see what characters a registry is permitting. Therefore, up to this point, we have been requesting, from every registry we enable IDN for, a single ordered list of all characters they permit, full stop. This enables us to see if there are any homographs; if there are, we can then further analyse their tables to see if they will be a problem in practice. I would suggest that ICANN make the production and publication of such a list a guideline in the same way that production of individual tables is. One is merely a reformatting and agglomeration of the other, so the additional work should not be great.

Guideline 4)

ETHIOPIC WORDSPACE (U+1361) is classified here as "necessary punctuation". This character is a homograph of : (colon), which is a "character with a well-established function as a protocol element" - a set later prohibited in the same guideline, for very good reasons.

Without knowing the exact use of the character in Ethiopic, I don't want to be dogmatic about this, but I'm concerned about it being classified explicitly by the guidelines as "necessary punctuation", and even about the existence of such a class as "necessary punctuation". ASCII domain names have coped for many years using only hyphen (and CamelCase, presentationally) as a separator. I think the real necessity of any punctuation, particularly that which spoofs protocol elements, needs extremely careful examination.

Additionally, I am also hoping that IETF processes will eventually lead to a revision of the IDN guidelines which use an inclusive approach, and which focus almost exclusively on letters and numbers in the various scripts. There is a danger that their view of "necessary punctuation" might not agree with the one in the guidelines.

Secondly, this guideline says that "such-and-such is not allowed", but then says any registry may make exceptions merely by documenting them. This rather removes any force the guideline may have had. While I know we can never see the future clearly, what sort of exceptions are envisaged, and why should we be allowing them, given the wiseness of prohibiting all the character classes explicitly listed in Guideline 4?

Thank you for the opportunity to comment on this draft; I hope you find my comments helpful, and I look forward to hearing any feedback you may have on them, and to seeing a further draft of the guidelines.

Gerv

[0] http://www.unicode.org/draft/reports/tr36/tr36.html#Security_Levels_and_Alerts

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

Comments and question on the proposed revision to the IDN guidelines

- *To*: idn-guidelines@xxxxxxxxx
- *Subject*: Comments and question on the proposed revision to the IDN guidelines
- *From*: Paul Hoffman <phoffman@xxxxxxxxxx>
- *Date*: Tue, 11 Oct 2005 15:17:51 -0700

The following are comments on the the proposal for "Guidelines for the Implementation of Internationalized Domain Names version 2". They are in response to the request for comments posted. These comments are a joint posting from Paul Hoffman and Bret Fausett.

1) The proposal would benefit greatly from a statement of purpose. Many parts of the proposal appear intended to deal with domain name spoofing below the TLD; however, the proposal fall very short of that goal by limiting itself to IDNs. If the proposal is intended to address domain name spoofing, a problem not limited to IDNs, ICANN should state whether it also intends to create new proposals about potential spoofing within the ASCII character set (such as "paypa1.com", "icannn.org", and "paypal.co"). If the proposal is not about name spoofing, a clear statement of its purpose is greatly needed.

2) The proposal is titled "Guidelines," but it contains many statements about what a registry must and must not do; the "Guidelines" are replete with words such as "will" and "must." ICANN should either change the title of the document or change the imperative verbs to ones that sound more like guidelines. In specific, the proposal never says what will happen if a particular TLD administrator does not follow the guidelines. Will ICANN re-assign the TLD to an administrator that agrees to follow the guidelines? If not, what possible penalties will be meted out for variance from the guidelines? For example, both MuseDoma and VeriSign contributed to the revisions for version 2, but both are out of compliance with both version 1 and version 2 of the guidelines. In addition, many large ccTLDs have ignored the requirements of version 1, and yet no action appears to have been taken against them. The "Guidelines" also refer to "top-level domain registries," a broad term which encompasses both gTLD and ccTLD registries. If the "Guidelines" were intended to apply to ccTLD registries, then the mandatory words such as "will" and "must," as described above, should be revised to more accurately reflect the role of ICANN in setting policy for ccTLDs.

3) Version 1 of the guidelines has been in existence for two years, and now ICANN is proposing to change it by adding many more rules. To the best of our knowledge, however, ICANN has not previously published a report on the successes and failures in version 1. From the very small number of TLDs represented at <http://www.iana.org/assignments/idn/registered.htm>, it appears that version 1 is thinly-implemented. The proposal for version 2 of the guidelines are significantly more restrictive than version 1. It seems unlikely that adding many new restrictions will make version 2 more successful than version 1. A description of both ICANN's and TLD registry operators' experience with Version 1 would provide necessary information to evaluate the appropriateness of the new rules now under consideration.

4) Not only do the proposed revisions not have a report about the success or failure of the version 1 guidelines, they provide no list of the changes between versions and why each was made. Giving the complete

list of additional restrictions and the logic behind them could help bring out more useful discussion.

5) Paragraph 6 of the proposed "Guidelines" purports to place a mandatory obligation on TLD registries to participate in a loose and undefined "collaboration" with other entities. While collaboration and consensus are laudable goals, the undefined obligation describes no forum in which such discussions should take place and no process by which to reach possible agreement on IDN implementation and registration practices. We recommend that ICANN create an appropriate cross-Supporting Organization forum, involving both the ccNSO and the gNSO, designed to facilitate the recommended collaboration.

Thank you for considering these comments.

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

Comments regarding the draft Guidelines for the Implementation of Internationalized Domain Names, v2.0

- *To*: idn-guidelines@xxxxxxxxx
- *Subject*: Comments regarding the draft Guidelines for the Implementation of Internationalized Domain Names, v2.0
- *From*: Neil Harris <neil.harris@xxxxxxxxxxxxxxxxxxx>
- *Date*: Tue, 11 Oct 2005 16:44:34 +0100

Dear ICANN members,

I am encouraged by the work being carried out on as part of the process of preparing the next generation of the Guidelines for the Implementation of Internationalized Domain Names, in particular with regard to the IDN homograph-spoofing and invalid characters issues. I feel that the multi-pronged approach being taken in the current draft has the potential to be effective, if pursued with sufficient vigor.

I would like to comment on the proposition that some character ranges should be blacklisted entirely. I approve wholeheartedly with this principle, as any reduction of the character repertoire will greatly ease the effort of looking for homographs.

I believe that a precautionary principle should be used, where many ranges of characters that are not useful for label creation should be blacklisted by default. I propose the following characters for blacklisting:

1: All ASCII characters other than letters, digits and HYPHEN-MINUS

Rationale: These characters are not allowed in RFC 1035, and their addition will serve no purpose for internationalization purposes. In addition, many of them have special meanings to computer programs, for example, as part of E-mail addresses, URLs, or as quote characters in various systems such as SQL queries or command-line interpreters. Disallowing these characters will reduce the possiblity of attacks on other protocols by their inclusion in domain names.

2: Any character which, after NAMEPREP processing, generates an ASCII character other than letters, digits or HYPHEN-MINUS

Rationale: these characters can potentially be expanded by software to ASCII punctuation characters during IDN processing and be passed through to lower-level DNS calls, allowing spoofing risks as outlined above. There are too many of these characters to list here.

In particular, these should include:

2a: Visual spoofs of SOLIDUS:

U+0337 COMBINING SHORT SOLIDUS OVERLAY
U+0338 COMBINING LONG SOLIDUS OVERLAY
U+2044 FRACTION SLASH
U+2215 DIVISION SLASH
U+23AE INTEGRAL EXTENSION

U+29F6 SOLIDUS WITH OVERBAR
U+29F8 BIG SOLIDUS
U+2AFB TRIPLE SOLIDUS BINARY RELATION
U+2AFD DOUBLE SOLIDUS OPERATOR
U+FF0F FULLWIDTH SOLIDUS
U+3033 VERTICAL KANA REPEAT MARK UPPER HALF

2b: Visual spoofs of FULL STOP and other label separators, which should never appear in a label:

U+2024 ONE DOT LEADER
U+2027 HYPHENATION POINT
U+06D4 ARABIC FULL STOP
U+0702 SYRIAC SUBLINEAR FULL STOP
U+3002 IDEOGRAPHIC FULL STOP
U+FF0E FULLWIDTH FULL STOP
U+FF61 HALFWIDTH IDEOGRAPHIC FULL STOP

3: Any character that is a visual spoof of any ASCII character other than letters or digits, or appears to contain a visual spoof of one of these characters, which is not detected by their presence in NAMEPREP output.

Rationale: Whilst these characters are not a danger to software, they can be used to create confusion in users, for example, by creating URLs that mislead users into thinking that they are visiting a different website to the one given.

A partial list of these characters is given at the end of this E-mail.

4: All characters labeled as Non-XID, Pattern_Syntax, IDN-Illegal or IDN-Deleted in http://www.unicode.org/reports/tr36/idn-chars.html

Rationale: these are either not useful for creating names or identifiers, or explicity violate IDN guidelines.

5: Spacing and filler characters, namely

U+0020 SPACE
U+00A0 NO-BREAK SPACE
U+115F HANGUL CHOSEONG FILLER
U+1160 HANGUL JUNGSEOUNG FILLER
U+2000 EN QUAD
U+2001 EM QUAD
U+2002 EN SPACE
U+2003 EM SPACE
U+2004 THREE-PER-EM SPACE
U+2005 FOUR-PER-EM SPACE
U+2006 SIX-PER-EM SPACE
U+2007 FIGURE SPACE
U+2008 PUNCTUATION SPACE
U+2009 THIN SPACE
U+200A HAIR SPACE
U+200B ZERO WIDTH SPACE
U+202F NARROW NO-BREAK SPACE
U+205F MEDIUM MATHEMATICAL SPACE
U+3000 IDEOGRAPHIC SPACE
U+3164 HANGUL FILLER
U+FEFF ZERO WIDTH NO-BREAK SPACE
U+FFA0 HALFWIDTH HANGUL FILLER

Rationale: Existing domain names cannot contain spaces. There is no practical reason for IDNs to contain spaces. In addition, some filler characters break some rendering engines, allowing to typograpic spoofing attacks.

6: Line separators

U+2028 LINE SEPARATOR

U+2029 PARAGRAPH SEPARATOR

Rationale: Neither of these can serve any practical purpose in a domain name.

7: Ideographic description characters, namely
U+2FF0 IDEOGRAPHIC DESCRIPTION CHARACTER LEFT TO RIGHT
U+2FF1 IDEOGRAPHIC DESCRIPTION CHARACTER ABOVE TO BELOW
U+2FF2 IDEOGRAPHIC DESCRIPTION CHARACTER LEFT TO MIDDLE AND RIGHT
U+2FF3 IDEOGRAPHIC DESCRIPTION CHARACTER ABOVE TO MIDDLE AND BELOW
U+2FF4 IDEOGRAPHIC DESCRIPTION CHARACTER FULL SURROUND
U+2FF5 IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM ABOVE
U+2FF6 IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM BELOW
U+2FF7 IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM LEFT
U+2FF8 IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM UPPER LEFT
U+2FF9 IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM UPPER RIGHT
U+2FFA IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM LOWER LEFT
U+2FFB IDEOGRAPHIC DESCRIPTION CHARACTER OVERLAID

Rationale: Unless CJK experts disagree, I cannot see these as being useful for IDNs, as
* most rendering engines cannot use them to render glyphs,
* if used as rendering hints, they allow multiple descriptions of the same ideograph, allowing multiple Unicode strings to map to the same visual appearance, which is a spoofing risk
* their use as printing characters, if not supported as rendering hints, is confusing, and again allows multiple representations of the same ideogram

8: All characters from all scripts described as "Ancient Scripts" in the Unicode code charts, such as Old Italic, Cuneiform, Old Persian, Ugaritic, Linear B Syllabary, Linear B Ideograms, Aegean Numbers, the Cypriot Syllabary, Gothic, and Runic.

Rationale: there appears to be no utility in creating names in dead languages.

9: Presentation forms from all scripts.

Rationale: as I understand it, presentation forms are only required for compatibility reasons, and provide alternative ways of representing the same characters that can be represented in other ways using standard Unicode characters. Even though some of these may be correctly dealt with by NAMEPREP, having more than one way of representing the same thing is generally a bad idea.

10: Musical symbols, currency symbols, specials, tags, layout controls, variation selectors.

Rationale: these characters seem unnecessary for use in names.

11: All private use characters, surrogates, and noncharacters.

Rationale: these characters are unusable in names.

I am aware that the ranges given here will probably overlap with one another, and may duplicate other lists of deprecated characters, such as those in UTR #36. Nevertheless, I believe that the precautionary principle suggests that these characters should be universally blacklisted in domain name labels, with specific exceptions if justified on a case-by-case basis for linguistic reasons.

I hope this input is useful as part of the consultation process.

Sincerely,

Neil Harris
Media Channel Limited

List of characters in class 3 above:

U+01C3 ; LATIN LETTER RETROFLEX CLICK ; -> EXCLAMATION MARK; [nameprep: LATIN LETTER RETROFLEX CLICK]
U+05C3 ; HEBREW PUNCTUATION SOF PASUQ ; -> COLON; [nameprep: HEBREW PUNCTUATION SOF PASUQ]
U+05F4 ; HEBREW PUNCTUATION GERSHAYIM ; -> QUOTATION MARK; [nameprep: HEBREW PUNCTUATION GERSHAYIM]
U+321D ; PARENTHESIZED KOREAN CHARACTER OJEON ; -> LEFT PARENTHESIS, HANGUL SYLLABLE O, HANGUL SYLLABLE JEON, RIGHT PARENTHESIS; [nameprep: ch321D]
U+321E ; PARENTHESIZED KOREAN CHARACTER O HU ; -> LEFT PARENTHESIS, HANGUL SYLLABLE O, HANGUL SYLLABLE HU, RIGHT PARENTHESIS; [nameprep: ch321E]
U+01C3 ; LATIN LETTER RETROFLEX CLICK ; -> EXCLAMATION MARK; [nameprep: LATIN LETTER RETROFLEX CLICK]
U+05C3 ; HEBREW PUNCTUATION SOF PASUQ ; -> COLON; [nameprep: HEBREW PUNCTUATION SOF PASUQ]
U+05F4 ; HEBREW PUNCTUATION GERSHAYIM ; -> QUOTATION MARK; [nameprep: HEBREW PUNCTUATION GERSHAYIM]
U+321D ; PARENTHESIZED KOREAN CHARACTER OJEON ; -> LEFT PARENTHESIS, HANGUL SYLLABLE O, HANGUL SYLLABLE JEON, RIGHT PARENTHESIS; [nameprep: ch321D]
U+321E ; PARENTHESIZED KOREAN CHARACTER O HU ; -> LEFT PARENTHESIS, HANGUL SYLLABLE O, HANGUL SYLLABLE HU, RIGHT PARENTHESIS; [nameprep: ch321E]
U+00BD ; VULGAR FRACTION ONE HALF ; -> LATIN SMALL LETTER L, SOLIDUS, DIGIT TWO; [nameprep: DIGIT ONE, FRACTION SLASH, DIGIT TWO]
U+01C3 ; LATIN LETTER RETROFLEX CLICK ; -> EXCLAMATION MARK; [nameprep: LATIN LETTER RETROFLEX CLICK]
U+2039 ; SINGLE LEFT-POINTING ANGLE QUOTATION MARK ; -> LESS-THAN SIGN; [nameprep: SINGLE LEFT-POINTING ANGLE QUOTATION MARK]
U+203A ; SINGLE RIGHT-POINTING ANGLE QUOTATION MARK ; -> GREATER-THAN SIGN; [nameprep: SINGLE RIGHT-POINTING ANGLE QUOTATION MARK]
U+2044 ; FRACTION SLASH ; -> SOLIDUS; [nameprep: FRACTION SLASH]
U+2154 ; VULGAR FRACTION TWO THIRDS ; -> DIGIT TWO, SOLIDUS, DIGIT THREE; [nameprep: DIGIT TWO, FRACTION SLASH, DIGIT THREE]
U+2155 ; VULGAR FRACTION ONE FIFTH ; -> LATIN SMALL LETTER L, SOLIDUS, DIGIT FIVE; [nameprep: DIGIT ONE, FRACTION SLASH, DIGIT FIVE]
U+2156 ; VULGAR FRACTION TWO FIFTHS ; -> DIGIT TWO, SOLIDUS, DIGIT FIVE; [nameprep: DIGIT TWO, FRACTION SLASH, DIGIT FIVE]
U+2159 ; VULGAR FRACTION ONE SIXTH ; -> LATIN SMALL LETTER L, SOLIDUS, CYRILLIC SMALL LETTER BE; [nameprep: DIGIT ONE, FRACTION SLASH, DIGIT SIX]
U+215A ; VULGAR FRACTION FIVE SIXTHS ; -> DIGIT FIVE, SOLIDUS, CYRILLIC SMALL LETTER BE; [nameprep: DIGIT FIVE, FRACTION SLASH, DIGIT SIX]
U+215B ; VULGAR FRACTION ONE EIGHTH ; -> LATIN SMALL LETTER L, SOLIDUS, GURMUKHI DIGIT FOUR; [nameprep: DIGIT ONE, FRACTION SLASH, DIGIT EIGHT]
U+2215 ; DIVISION SLASH ; -> SOLIDUS; [nameprep: DIVISION SLASH]
U+3015 ; RIGHT TORTOISE SHELL BRACKET ; -> RIGHT SQUARE BRACKET; [nameprep: RIGHT TORTOISE SHELL BRACKET]
U+321D ; PARENTHESIZED KOREAN CHARACTER OJEON ; -> LEFT PARENTHESIS, HANGUL SYLLABLE O, HANGUL SYLLABLE JEON, RIGHT PARENTHESIS; [nameprep: ch321D]
U+321E ; PARENTHESIZED KOREAN CHARACTER O HU ; -> LEFT PARENTHESIS, HANGUL SYLLABLE O, HANGUL SYLLABLE HU, RIGHT PARENTHESIS; [nameprep: ch321E]
U+33AE ; SQUARE RAD OVER S ; -> LATIN SMALL LETTER R, LATIN SMALL LETTER A, LATIN SMALL LETTER D, SOLIDUS, LATIN SMALL LETTER TONE FIVE; [nameprep: LATIN SMALL LETTER R, LATIN SMALL LETTER A, LATIN SMALL LETTER D, DIVISION SLASH, LATIN SMALL LETTER S]
U+33AF ; SQUARE RAD OVER S SQUARED ; -> LATIN SMALL LETTER R, LATIN SMALL LETTER A, LATIN SMALL LETTER D, SOLIDUS, LATIN SMALL LETTER TONE FIVE, DIGIT TWO; [nameprep: LATIN SMALL LETTER R, LATIN SMALL LETTER A, LATIN SMALL LETTER D, DIVISION SLASH, LATIN SMALL LETTER S, DIGIT TWO]
U+FE14 ; PRESENTATION FORM FOR VERTICAL SEMICOLON ; -> SEMICOLON; [nameprep: chFE14]
U+FE15 ; PRESENTATION FORM FOR VERTICAL EXCLAMATION MARK ; -> EXCLAMATION MARK; [nameprep: chFE15]
U+FE3F ; PRESENTATION FORM FOR VERTICAL LEFT ANGLE BRACKET ; -> CIRCUMFLEX ACCENT; [nameprep: LEFT ANGLE BRACKET]
U+FE5E ; SMALL RIGHT TORTOISE SHELL BRACKET ; -> RIGHT SQUARE BRACKET; [nameprep: RIGHT TORTOISE SHELL BRACKET]
U+00BC ; VULGAR FRACTION ONE QUARTER ; -> DIGIT ONE, SOLIDUS, CHEROKEE LETTER SE; [nameprep: DIGIT ONE, FRACTION SLASH, DIGIT FOUR]
U+00BD ; VULGAR FRACTION ONE HALF ; -> DIGIT ONE, SOLIDUS, DIGIT TWO; [nameprep: DIGIT ONE, FRACTION SLASH, DIGIT TWO]
U+01C3 ; LATIN LETTER RETROFLEX CLICK ; -> EXCLAMATION MARK; [nameprep: LATIN LETTER RETROFLEX CLICK]

U+2039 ; SINGLE LEFT-POINTING ANGLE QUOTATION MARK ; -> LESS-THAN SIGN; [nameprep: SINGLE LEFT-POINTING ANGLE QUOTATION MARK]
U+203A ; SINGLE RIGHT-POINTING ANGLE QUOTATION MARK ; -> GREATER-THAN SIGN; [nameprep: SINGLE RIGHT-POINTING ANGLE QUOTATION MARK]
U+2044 ; FRACTION SLASH ; -> SOLIDUS; [nameprep: FRACTION SLASH]
U+2154 ; VULGAR FRACTION TWO THIRDS ; -> DIGIT TWO, SOLIDUS, DIGIT THREE; [nameprep: DIGIT TWO, FRACTION SLASH, DIGIT THREE]
U+2155 ; VULGAR FRACTION ONE FIFTH ; -> DIGIT ONE, SOLIDUS, DIGIT FIVE; [nameprep: DIGIT ONE, FRACTION SLASH, DIGIT FIVE]
U+2156 ; VULGAR FRACTION TWO FIFTHS ; -> DIGIT TWO, SOLIDUS, DIGIT FIVE; [nameprep: DIGIT TWO, FRACTION SLASH, DIGIT FIVE]
U+215A ; VULGAR FRACTION FIVE SIXTHS ; -> DIGIT FIVE, SOLIDUS, CYRILLIC SMALL LETTER BE; [nameprep: DIGIT FIVE, FRACTION SLASH, DIGIT SIX]
U+215F ; FRACTION NUMERATOR ONE ; -> DIGIT ONE, SOLIDUS; [nameprep: DIGIT ONE, FRACTION SLASH]
U+2215 ; DIVISION SLASH ; -> SOLIDUS; [nameprep: DIVISION SLASH]
U+3015 ; RIGHT TORTOISE SHELL BRACKET ; -> RIGHT SQUARE BRACKET; [nameprep: RIGHT TORTOISE SHELL BRACKET]
U+321D ; PARENTHESIZED KOREAN CHARACTER OJEON ; -> LEFT PARENTHESIS, HANGUL SYLLABLE O, HANGUL SYLLABLE JEON, RIGHT PARENTHESIS; [nameprep: ch321D]
U+321E ; PARENTHESIZED KOREAN CHARACTER O HU ; -> LEFT PARENTHESIS, HANGUL SYLLABLE O, HANGUL SYLLABLE HU, RIGHT PARENTHESIS; [nameprep: ch321E]
U+33AE ; SQUARE RAD OVER S ; -> LATIN SMALL LETTER R, LATIN SMALL LETTER A, LATIN SMALL LETTER D, SOLIDUS, LATIN SMALL LETTER TONE FIVE; [nameprep: LATIN SMALL LETTER R, LATIN SMALL LETTER A, LATIN SMALL LETTER D, DIVISION SLASH, LATIN SMALL LETTER S]
U+33AF ; SQUARE RAD OVER S SQUARED ; -> LATIN SMALL LETTER R, LATIN SMALL LETTER A, LATIN SMALL LETTER D, SOLIDUS, LATIN SMALL LETTER TONE FIVE, DIGIT TWO; [nameprep: LATIN SMALL LETTER R, LATIN SMALL LETTER A, LATIN SMALL LETTER D, DIVISION SLASH, LATIN SMALL LETTER S, DIGIT TWO]
U+33C6 ; SQUARE C OVER KG ; -> CHEROKEE LETTER TLI, SOLIDUS, LATIN SMALL LETTER K, LATIN SMALL LETTER G; [nameprep: LATIN SMALL LETTER C, DIVISION SLASH, LATIN SMALL LETTER K, LATIN SMALL LETTER G]
U+33DF ; SQUARE A OVER M ; -> CANADIAN SYLLABICS CARRIER GHO, SOLIDUS, CANADIAN SYLLABICS CARRIER GO; [nameprep: ch33DF]
U+FE14 ; PRESENTATION FORM FOR VERTICAL SEMICOLON ; -> SEMICOLON; [nameprep: chFE14]
U+FE15 ; PRESENTATION FORM FOR VERTICAL EXCLAMATION MARK ; -> EXCLAMATION MARK; [nameprep: chFE15]
U+FE3F ; PRESENTATION FORM FOR VERTICAL LEFT ANGLE BRACKET ; -> CIRCUMFLEX ACCENT; [nameprep: LEFT ANGLE BRACKET]
U+FE5E ; SMALL RIGHT TORTOISE SHELL BRACKET ; -> RIGHT SQUARE BRACKET; [nameprep: RIGHT TORTOISE SHELL BRACKET]

and, for good measure,

Missing symmetry pairs from the above, just in case
U+3014 LEFT TORTOISE SHELL BRACKET
U+FE5D SMALL LEFT TORTOISE SHELL BRACKET

<http://www.unicode.org/reports/tr31/tr31-5.html>

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

IAB comments on ICANN IDN Guidelines

- *To*: idn-guidelines@xxxxxxxxx
- *Subject*: IAB comments on ICANN IDN Guidelines
- *From*: Leslie Daigle <leslie@xxxxxxxxxxxxxxx>
- *Date*: Fri, 14 Oct 2005 15:20:52 -0400

The IAB has reviewed the IDN Guidelines that ICANN has prepared for open input, published Sept 20, 2005 that can be found on the following URI:

http://icann.org/announcements/announcement-20sep05.htm

We have the following comments:

In the proposed guidelines, we see the following text:

> 3. (a) In implementing the IDN standards, top-level domain registries will associate each label in a registered internationalized domain name, as it appears in their registry, with a single language or a single script using accepted designators for both. The restriction, in either case, is intended to limit the set of permitted characters within a label. If greater specificity is desired, the association may be made by combining both a language designator and a script designator. Alternatively, a label may be associated with a set of languages, or with more than one designator under the conditions described below. Language designators are illustrated in RFC 3066 (http://www.rfc-editor.org/ rfc/rfc3066.txt).

RFC 3066 is currently under review by the IETF, and the registry it defines will always be updated. Because of this, the IAB believes it is problematic to explicitly refer to RFC 3066, but would instead be appropriate say the following:

> Language designators or "language tags" are associated with an IANA registry, currently maintained at http://www.iana.org/assignments/ language-tags and registered according to rules set out in BCP 47. Registries should consult the registry and BCP 47.

Best regards,

Leslie Daigle,
Chair, Internet Architecture Board.

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

INTA IDN Response

- *To*: <idn-guidelines@xxxxxxxxx>
- *Subject*: INTA IDN Response
- *From*: "Michael Heltzer" <mheltzer@xxxxxxxx>
- *Date*: Tue, 18 Oct 2005 11:22:50 -0400

A. INTA Is the World's Leading Voice on Trademark Law

The International Trademark Association ("INTA") is a 127 year-old global organization with members in over 180 countries. One of INTA's key goals is the promotion and protection of trademarks as a primary means for consumers to make informed choices regarding the products and services they purchase.

For the last eleven years, INTA has also been the leading voice of trademark owners within the Internet community, serving as a founding member of the Intellectual Property Constituency of the Internet Corporation for Assigned Names and Numbers ("ICANN").

B. Purpose of these Submissions

ICANN has opened a 30-day public comment period on the revised version of the Guidelines for the Implementation of Internationalized Domain Names ("IDN Guidelines"). INTA is filing these submissions in response to ICANN's request for feedback.

C. Summary

The IDN Guidelines deal with the implementation of Internationalized Domain Names ("IDNs"). The IDN Guidelines do not mention, let alone deal with, any Whois issues that will arise from the implemenation of IDNs.

INTA respectfully submits that the draft IDN Guidelines must deal with the creation, maintenance and publication of Whois data, and that a further draft dealing with Whois issues should be prepared and resubmitted for public comment.

D. Implementation of IDNs

The central goal of the proposed IDN Guidelines is to enable end users to view IDNs, without altering the DNS protocols.

The IDN Guidelines seek to protect the security of the Internet by preventing users who enter a single IDN from being connected to different servers based on different interpretations of that domain name. *See* RFC 3490 at § 10.0.

As ICANN is aware, the increased availability of characters created by IDNs presents additional opportunities for homograph domain name spoofing and URI spoofing. *See* "ICANN Statement on IDN Homograph Attacks and Request for Public Comment," dated February 23, 2005.

Although the IDN Guidelines will curtail the opportunity for abusive registration of IDNs, abuses will occur. When that happens, trademark owners and others must be able to determine who is responsible in order to seek redress and prevent further infringement and public deception. The establishment of Whois guidelines therefore goes hand in hand with, and should be a part of, the IDN Guidelines.

E. History of the Whois Database

The Whois database refers to the publicly available online system that provides access to ownership and contact information regarding domain name registrations.

Under Section 3.3 of the Registrar Accreditation Agreement of ICANN registrars in the familiar .com, .net, .org, and other generic top level domains ("gtld's") must provide a free, publicly accessible online database of domain name registrant information, providing at least the following:

- the registered domain name;
- the registrant's name and postal address;
- the administrative contact's and the technical contact's names, postal addresses, email addresses, telephone numbers and (if available) fax numbers;

- original registration date and expiration date;
- the registrar's name; and
- the names of the primary and secondary nameservers.

The public availability of domain name ownership information is consistent with how the Whois database has always been operated from the very early days of the Internet. Indeed, both RFC 812 (March 1, 1982) and its replacement, RFC 954 (October 1985), refer to the Whois database as "a netwide directory service for Internet [ARPANET] users."

With the growth of the use of the Internet, and in particular commercial use of the World Wide Web, the Whois database has become an important source of information for contacting the person or persons responsible for administering domain names (including domain name holders) for public policy purposes. For instance, the Whois database is widely used as an identification tool in (i) law enforcement for identifying counterfeiters and combating other fraud, (ii) consumer protection by allowing the ready identification of online merchants, and (iii) protection of intellectual property rights by helping to identify infringers.

F. Whois Concerns with respect to IDNs

The NRC publication entitled *Signposts in Cyberspace: The Domain Name System and Internet Navigation*, identified three key Whois issues with respect to IDNs, which should be resolved at the same time as IDN implementation issues are being considered. These Whois issues are as follows:

1. What characters should be acceptable in a Whois query? The choices include not only Unicode, but also IDNA puny code and local characters, or some combination of them.

2. What language should be acceptable in a response, and how should it be encoded? The choices of language include the language of the nation in which the registrar or the registrant is located, or any "international language" (such as English, Chinese, French Spanish, Russian, Arabic, etc.) or one designated language (such as English). If the language is other than English, coding issues will arise.

3. Since IDN practices for complex languages create packages of reserved names (containing variant characters), how much information should Whois provide about

other names in the package in response to a query about one of them? Apparently, examples have been shown of some Chinese labels that could generate hundreds of variants.

National Research Council of the National Academies, National Academies Press (2005, prepublication issue, Section 5.7.2 (titled “Whois and Internationalized Domain Names”), p. 5-62 et seq.

G. Conclusion

INTA respectfully submits that ICANN should prepare a further draft of the proposed IDN Guidelines, dealing with the Whois issues identifed above, and any other Whois issues of concern, and then publish the revised draft for public comment.

INTA appreciates the opportunity to make these submissions in response to ICANN’s request for feedback regarding the proposed IDN Guidelines. INTA would also appreciate the opportunity to be involved in drafting the revised IDN Guidelines with respect to the Whois issues identified above, or in consulting with ICANN with respect to those issues.

Michael E. Heltzer

External Relations Manager

International Trademark Association

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

Comments on ICANN IDN Guidelines V 2.0

- *To*: idn-guidelines@xxxxxxxxx
- *Subject*: Comments on ICANN IDN Guidelines V 2.0
- *From*: Abdulaziz Al-Zoman <zoman@xxxxxxxxxx>
- *Date*: Wed, 19 Oct 2005 14:30:56 +0300

Dear ICANN,

The following are comments as a response to the request for comments on the âICANN Guidelines for the Implementation of Internationalized Domain Names, Draft v2.0â.

These comments are from the Steering Committee of the Arabic Domain Names Pilot Project (www.arabic-domains.org).

* Unfortunately, the ICANN IDN Guidelines V.2 is still working on a handicapped IDN solution (i.e., ML.English) that does not support full IDN on a TLD level (i.e., ML.ML).

* The ICANN IDN Guidelines V.2 reflects the experiences of the IDN registries who have implemented version 1.0 (i.e., registries which provide ML.English, such as VerSign). This excludes the experiences collected by different entities around the world who strive to support their languages on domain names.

* The current IDN implementations as suggested by the ICANN IDN Guidelines V. 1.0 and 2.0 (i.e., ML.English) still are not suitable for languages that are not Latin-based, for example, languages written from right-to-left (e.g., Arabic, Farsi, Urdu) or ideographic languages (e.g., Chinese, Japanese, Korean).

* A more practical approach even for testing proposes is to start the IDN support at a ccTLD level rather than on a gTLD. So that the TLD is written in a specific language (e.g. Arabic) that will be supported on the SLD controlled by a character set table. In this case characters from other scripts (e.g., Farsi, Urdu, â) will not be confused with visually confusable characters.

* It is strongly believed that concerns and issues that are raised by the guidelines regarding IDN implementations would be addressed when internationalizing ccTLDs are supported.

* The Arabic domain names pilot project (www.arabic-domains.org) supports the following principles that have been stated in the proposal submitted by the Chinese Domain Name Consortium (CDNC) (www.icann.org/announcements/idn-tld-cdnc.pdf) to ICANN, namely:

  - Give the priority to internationalizing ccTLDs. To ensure the system stability, itâs recommended to internationalize

ccTLD before internationalize gTLD.

- For convenience purpose, only one form of language character variant of internationalized ccTLD is accepted. Considering that some countries or regions may have character variants, only one form of character sets shall be chosen for IDN use by each sponsored registry.

- Supported by their own governments, ccTLD registries or authorized agencies make their own choice of which IDN character sets for their ccTLDs.

- Register and operate the internationalized ccTLDs in the root DNS server in the form of IDNA Punycode.

Thank you for giving us the chance to express our opinion.

Best regards,

On behave of the Arabic Domain Names Pilot Project

Dr. Abdulaziz H. Al-Zoman
Chairman of the Steering Committee (AND Pilot Project), and
Director of SaudiNIC (.sa ccTLD Registry)
zoman@xxxxxxxxxx

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

Comments on 20 September draft revised IDN guidelines

- *To*: idn-guidelines@xxxxxxxxx
- *Subject*: Comments on 20 September draft revised IDN guidelines
- *From*: John C Klensin <klensin@xxxxxxx>
- *Date*: Sat, 22 Oct 2005 12:25:47 -0400

Despite considerable interest in issues associated with IDNs, have refrained from commenting on this draft until now for three reasons. First, I have wanted to understand how other comments were running. Second, and more important, I do not believe that the current draft for version 2.0 represents the right direction for the Internet community and have been trying to figure out how to address that issue. Third, I have gradually lost most of the confidence I once had in the quality of ICANN's review and approval processes, so I am not sure that writing these notes is worth the trouble.

So, while I applaud the efforts of the members of the Guidelines committee to put this draft together and of ICANN in finally initiating some real work on IDNs after years of promises, I am deeply concerned about this draft and its contents.

Comments on "Guidelines for the Implementation of Internationalized Domain Names, Draft Version 2.0", dated 20 September 2005.

(1) Parts of the draft have the wrong tone and may represent the wrong approach.

It is clearly in the best interest of the Internet community that all domains -- gTLDs, ccTLDs, and domains at the second level and beyond -- have consistent policies with regard to IDNs. If we do not have consistent policies, then users do not know what to expect.

If policies are not consistent, reasonable protection for users encourages --some would say "forces"-- applications implementers to develop per-domain profiles or lists of domains with acceptable or unacceptable behavior. We have, of course, already seen that response. Per-domain profiles, different IDNA tables used in different applications, and similar behavior, no matter how well-intentioned (or even necessary) further damages user ability to predict how a particular program will behave and reduces compliance with the relevant standards. Full compliance with IDNA requires that applications display IDNs in fonts and glyphs appropriate to the relevant characters if it is possible to do so.

If a user obtains an application or operating system that is fully Unicode-capable and that has a full set of fonts, that user should, given the standard, be able to expect that he or she will never see a "punycode"-encoded name. Certainly such names should not be seen on an arbitrary basis or on a basis that is linked to particular domain names at the top levels of

the tree. But that is exactly where we have found ourselves with web browsers as a result of the recent IDN-"phishing" problems and ICANN's unwillingness to address IDN policy and permitted character issues when they were first identified almost five years ago, while there was still time to address them in the context of the basic design of IDNA. This draft does not significantly improve on the present situation.

In this area, and others, ICANN must decide what business it is in and how it expects to generate and, if appropriate enforce, policies. Where consistency across domains is important, as it is in this case, ICANN can, in principle, make policies applicable to the entire Internet community, insisting that conformance is a requirement of the long-standing provisions of RFC 1591 that all domain managers act as trustees for the global Internet community and act according to their responsibilities to that community. ICANN would also need to insist that domain managers carry out their obligations under the "recursive application" rule of RFC 1591, i.e., that they enforce requirements placed on them on all subdomain registrations they permit.

The existence of that authority in principle is, however, meaningless without the will and ability to apply it, and apply it consistently, in practice. Based on the history of the last few years, including ICANN's interactions with other organizations and governmental entities, it seems unlikely that ICANN, in practice, has any authority in this area that can be exercised in a sufficiently global way to provide users with a consistent and safe experience.

ICANN has an alternative which, in practice, would be more likely to serve the overall community well. That alternative is to carefully explain the issues, provide "best practices" guidelines for dealing with them and persuasive explanations of why those practices are appropriate and necessary, and clearly and logically identify the institutions that should have responsibility for various actions and controls. If that is done, then ICANN should be able to assert only the level of authority that it actually has in practice. It would then step back, in the presence of clear explanations of issues and alternatives, rely for enforcement on the good sense of domain administrators and managers, the workings of the marketplace, and the various governmental and judicial systems around the world.

To accomplish that end,

(a) Any document such as this one must clearly differentiate between requirements of the IDNA standard and recommendations or requirements imposed by ICANN or based on other community consensus.

(b) Any document such as this one should avoid stating requirements in terms of "ICANN commands and everyone else will comply". That type of construction was one of the reasons why some of the provisions of the initial version of the Guidelines were widely ignored even when they were sensible. Instead, ICANN should state a recommendation, explain why that recommendation is important, and, ideally, explain the adverse consequences --in terms of Internet behavior as seen by registrants or users-- of not following it.

(c) Documents such as this one should drop the pretense that gTLDs and ccTLDs are different from the standpoint of the Guidelines (the order in which IDNs should be deployed in different domains is a different issue). If ICANN has no practical enforcement capability in one case and no will to attempt to enforce policies in the other, there are no practical differences. Worse, making distinctions between "registries that have agreements with ICANN" and those which do not, and then imposing additional requirements on the former,

represents poor strategic policy as it discourages those registries for whom reaching agreements with ICANN is voluntary from ever reaching such agreements.

---

(2) Any set of rules or guidelines should make the locus of responsibility for specific implementations of the rules clear; this document does not do so.

As discussed above, this draft is laden with language about what registries "will" or "must" do. Independent of where the authority to make or enforce such statements comes from, it is important to identify the reasons for those choices, rather than possible alternatives, better than these guidelines do. Even more important, there is an industry practice of passing all responsibility for problematic registrations from registry to registrar to registrant. This is reasonable for, e.g., trademark conflicts where it may be plausible to expect the would-be registrant to take responsibility for determining rights in the chosen name. However, there are IDN issues involving name conflicts and name similarities in which only the registry, by inspecting its own databases, can determine whether it is appropriate to register a name. For traditional, LDH, domain name labels, registry-level appropriateness typically involves only a determination of whether the label is already present. For IDNs, the necessary determination may involve understanding whether a visually-confusable label exists, or whether a label is not permitted due to an existing label group (variant set). If registries fail to establish and enforce effective conventions in those areas, and harm results, the responsibility must rest largely with the registry.

---

(3) The draft does not go far enough to be significantly useful.

Despite being stated as strong requirements, paragraphs 3 and 4 of the Guidelines do not provide any real guidance for marginal cases. Any rules of this type should start from a clear statement of the principle that the use of the DNS as a source of precise and unique identifiers for Internet objects is paramount. Without that as a primary principle, the network as users know it simply ceases to function. ICANN, and all domain managers, need to accept and understand this principle and understand that it may force banning the use of some strings as IDNs even if those string would be culturally and linguistically reasonable in some language considered by itself. As a trivial example for English-language strings, the use of space characters, commas, and periods is usually necessary to form sentences or phrases. Yet those characters have always been prohibited as part of domain name labels to be used in applications because they would cause too much confusion and too many risks to the integrity of DNS references. Similarly strong rules should be applied to the use of such characters, or any character that maps onto them, in IDNs, ideally with no exceptions. If, counter to whatever guidelines or "best practice" statements exist, registries make exceptions to such rules, they must bear complete responsibility for any negative consequences.

Beyond that principle, Sections 3 and 4 indicate what code points may or may not be permissible by broad examples. That approach does not provide much guidance except to experts. While there may be many experts on a single language, there are few experts across all of the languages and scripts of the world. The approach of making specific prohibitions mentioned by Neil Harris in his posting about restrictions on 11 October appears to be a much more satisfactory method of dealing with these issues, and a better way to provide useful guidance, than by citing a few examples.

---

(4) The draft focuses on the registration process, rather than on impact on actual implementations and users and user experiences.

As I have indicated in other contexts, users do not generally use domain names. They use URLs or other URIs or IRIs, they use email addresses, and they may use other identifiers that incorporate domain names. The draft Guidelines identify one aspect of this issue in Section 5 but indicate only that the registry should "include in its documentation a description of the factors that determine the way that sequence appears at the user interface". I have no idea what that means; I would predict that the typical registry would not have a much better idea. That is not "guidance".

But, more generally, DNS registries typically deal with the registration of single labels at a single level of the system. Several confusing situations can be introduced by sequences of labels, especially sequences in different scripts. However, suggestions to restrict the language or script of labels at one level of the DNS tree based on the language or script used at another level have generally proven infeasible due to other DNS constraints. The Guidelines, especially if they are to be titled as "Guidelines for Implementation" should either clearly address these issues or should avoid them entirely, pointing if appropriate to other documents and efforts.

---

(5) The draft is internally confusing and creates new ambiguities.

The draft represents an odd mix of standards from different bodies, partially-ratified but still-changing technical reports, and other documents as reference sources. If the reader is left to interpret the intersections among, and applicability of, these documents, the only certain results are inconsistency and confusion.

In particular, the language and script registration rules of Section 3 rely on a mixture of an IETF Standard for language identification that was designed primarily for another purpose and may not be completely suitable for this one (RFC 3066), an ISO Standard for script identification (IS 15924) that has been controversial in some quarters and that does not contain guidelines for use in this type of context, and a Unicode character properties list (UTR 23) that evolves as characters are added to Unicode and needs change. No guidance is given about how those various standards can and should be used together. An IETF effort (the products of the LTRU working group) that might provide some assistance in the area is not referenced (although it is mentioned under "Additional remarks", see below). Whether it should be depended upon now is questionable since it is not clear at what granularity it should be applied to this work but, since it is intended to supersede RFC 3066, ignoring it entirely seems inappropriate. The text indirectly indicates an understanding of the issues by indicating that the various standards "illustrate" the relevant designation, but, again, that approach provides little real guidance.

Worse, if the statements made in that paragraph are taken as rules or guidance, they would essentially prohibit the use of a few Indian languages, and a large number of African languages, in IDNs despite the fact that all of the required characters appear in the Unicode code tables. If there is a language or script for which Unicode encodings for some characters do not exist, it is probably appropriate and necessary, although painful, for ICANN to take the position that the encodings be

registered first and then that IETF extend the mappings permitted by IDNA before IDN registrations are permitted: there does not appear to be any stable alternative. However, if Unicode codings are available for all of the characters relevant to a particular language and script, but no standardized names exist for those characters as a single script or unique to that language, it seems unreasonable to ban, or even significantly postpone, IDN registrations for that language.

---

(6) The status of the "Additional remarks" is unclear.

Are these remarks part of the Guidelines? Suggestions about areas for future revisions? An indication that the Guidelines are complete enough for community comment but not for any instantiation into policy? Neither the text nor the "Additional remarks" title provides any answers to those questions. The questions are important because several of the remarks that appear to be statements of fact are actually statements of highly controversial opinions. For example, UTR 36 remains controversial both within and outside the Unicode development community.

Whether the status of ISO 639-3 (which is an improper way to refer to it, it is ISO/DIS 639-3) is "advanced stage" is in the mind of the beholder: the relevant technical committee identifies a "Publication target date" of "2006-12-03" and a "Status" of "Under Development". It is worth noting that the initial DIS version of ISO 10646, the ISO counterpart to the Unicode standard, was completely replaced by a new version and model at a later stage than 636-3 has now achieved.

There are other examples, but these comments are already too long. If the "Additional remarks" section is intended to suggest that this draft of new Guidelines is unsuitable for incorporation into any policy process as it is now written, I would completely agree. If that section somehow is not expected to count, then it is even more clear to me that the draft Guidelines need an extensive reworking, starting from different principles about policies, relationships, and actions and then continuing on to address the technical issues in a way that provides more actual guidance.

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

Unicode Security Subcommittee comments on ICANN Draft Revised IDN Guidelines v2.0

- *To*: idn-guidelines@xxxxxxxxx
- *Subject*: Unicode Security Subcommittee comments on ICANN Draft Revised IDN Guidelines v2.0
- *From*: Mark Davis <mark.davis@xxxxxxxxxxxxxxx>
- *Date*: Sat, 22 Oct 2005 09:19:34 -0700

The Unicode security subcommittee welcomes the progress that is being made on the ICANN "Guidelines for the Implementation of Internationalized Domain Names" (http://icann.org/general/idn-guidelines-20sep05.htm). There have been important improvements in the text that should significantly contribute to the security of IDN.

However, there are a number of areas where the text needs further improvement. First, it should be better coordinated with the work that the Unicode Consortium is doing in Unicode Technical Report #36: Unicode Security Considerations (http://www.unicode.org/reports/tr36/) and Unicode Technical Standard #39: Unicode Security Mechanisms (http://www.unicode.org/reports/tr39/). In addition, the current draft of the ICANN guidelines is progressing towards a script-based, rather than language-based approach to character repertoires. This is going in the right direction, but needs to go further, as outlined in our comment 1. The proposed guidelines are also better defined in their approach to punctuation and other symbols, but also do not go far enough; see our comment 3.

Our proposals for changes in the text are listed below.

1. In Clause 3, replace the use of "language" with "script" as an appropriate designator. Then allow for restrictions of the character repertoire within the script. In concert with this, remove the text at the end: "Unless there is need to associate individual labels in an IDN with different scripts, even where script-based policies are otherwise applied, ...to the new situation." and remove the 2nd paragraph of the additional remarks, since it is no longer relevant.

Rationale: Clause 3 still refers to languages as an acceptable designator to determine a character set suitable for IDN registration. The difficulty in using language for such purpose is exposed in Annex G of the Unicode Technical Report #36: http://www.unicode.org/reports/tr36/tr36-4.html#Language_Based_Security.

2. Replace "Unicode Technical Report #23 (http://www.unicode.org/erports/tr23)" by "Unicode Standard Annex #24: Script Names (http://www.unicode.org/reports/tr24";, and replace "UTR#23" by "UAX #24"

Rationale: From context this simply appears citing the wrong document (and misspelling the URL). Note that the two documents have different status: #24 is a UAX, whereas #23 is a UTR.

3. Replace "Exception to this is permissible for languages with established orthographies and conventions that require the commingled use of multiple scripts. Visually confusable characters from different scripts must not appear in a single label unless there are overriding legitimate linguistic reasons for doing so." by references to the visual confusability tests in UTS #39.

Rationale: As it stands, these subclauses are essentially impossible to test for in practice. A registry cannot hand-examine all registration proposals; there must be a clear, mechanical test for validity.

4. Reverse the sense of Clause 4.

Rationale: Clause 4 in many aspect weakens the punctuation restriction from the previous version of the guidelines as it move the text from informative notes to a more prescriptive guideline status. In that aspect it is a regrettable regression from the previous ICANN guidelines. It also does not endorse the "inclusion-based" approach requested in Clause 2. Again, the recommendation from UTR#36 should be followed in removing all punctuation not related to a specific script usage, while preserving a well documented set of very limited exceptions. The data files and mechanisms are specified in UTS #39.

5. In "*Additional remarks", replace "The deceptive use of visually confusable characters from different scripts is discussed in detail in the Unicode Technical Report #36 on âUnicode Security Conditionsâ at http://www.unicode.org/reports/tr36/";

with

"The deceptive use of visually confusable characters from different scripts is discussed in detail in the Unicode Technical Report #36 on âUnicode Security Considerationsâ at http://www.unicode.org/reports/tr36/. Limitations to the character repertoire available for IDNs are provided in tables presented under the heading âData filesâ in Unicode Technical Standard #39: Unicode Security Mechanisms (http://www.unicode.org/reports/tr39/)"

Rationale: The Unicode Consortium has split the security related material into a Technical Report, describing the issues (UTR #36) and a Unicode Technical Standard (UTS #39) which gives precise specifications and data files.

6. The text should be reorganized for clarity. For example, both clause 3a and 7 deal with publishing tables; they should be in the same section; the 3rd paragraph of "*Additional remarks"* does not make much sense as currently written. At minimum it should be split by concept and developed in separate paragraphs (script consideration, idn.idn, etc...)

Mark Davis
Chair, Unicode Security Subcommittee
President, Unicode Consortium

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

Comments on Guideline 4 in the v2.0 draft

- *To*: idn-guidelines@xxxxxxxxx
- *Subject*: Comments on Guideline 4 in the v2.0 draft
- *From*: Cary Karp <ck@xxxxxxxxxx>
- *Date*: Sat, 22 Oct 2005 19:21:52 +0200 (CEST)

Point 4 in the draft guidelines was the focus of considerable discussion by the members of the guideline task force, who also expected it to generate significant public commentary. Useful remarks have already appeared on the present forum and in communications via other channels. Additional commentary is likely to posted here as the deadline approaches, and people who are preparing to do so may find it worth noting a few background details about some of the wording in the draft. (This is a personal contribution to the discussion, expressing opinions that may be solely my own, and is not being made on behalf of the task force.)

The passage on which I am commenting is:

"Permissible code points will not include ... punctuation characters that lack grammatical significance in the language with which the IDN registration is associated (with necessary punctuation including characters such as the ETHIOPIC WORDSPACE in Amharic and the MIDDLE DOT in Catalan)..."

This wording was arrived at in real-time dialog with participants in the meeting on 'Unicode and IDN in Africa' (http://www.next.sn/unicode-idn-africa.html), which took place in Dakar at the same time the guidelines task force was meeting. I was on the one end of an IM link between the two venues (with Michael Everson on the other) but cannot describe the discussion on the African side. Suffice it to say that my immediate correspondent made it clear that the credibility of the guidelines in African contexts could prove highly dependent on explicit reference to the ETHIOPIC WORDSPACE in precisely the form that subsequently appeared in the draft guidelines. The further reference to the MIDDLE DOT was considered in the same dialog, prompted by the IDN requirements of the new .cat TLD, and similarly put forward as an absolute necessity.

As anticipated, concern was expressed about this on the public forum. Quoting Gervase Markham:

> ETHIOPIC WORDSPACE (U+1361) is classified here as 'necessary punctuation'. This character is a homograph of : (colon), which is a 'character with a well-established function as a protocol element' - a set later prohibited in the same guideline, for very good reasons. Without knowing the exact use of the character in Ethiopic, I don't want to be dogmatic about this, I'm concerned about it being classified explicitly by the guidelines as 'necessary punctuation', and even about the existence of such a class as 'necessary punctuation'. ASCII domain names have coped for many years using only hyphen (and CamelCase, presentationally) as a separator. I think the real necessity of any punctuation, particularly that which spoofs protocol elements, needs extremely careful examination."

The terminology used in the guidelines clearly needs further clarification and some concepts may be relabeled entirely. The term 'grammatical significance' will likely not be retained, but notions of necessary punctuation -- however they may ultimately be worded -- figure prominently in IDN policy.

The extent to which one language can comfortably be represented using a single punctuation mark says absolutely nothing about the number of such characters needed

for the comparably adequate representation of another. One of the most important requirements placed on the guidelines is keeping them reasonably immune to accusations of reflecting cultural bias (a goal that other comments suggest they have yet to attain). Statements to the effect of, 'what is sufficient for anglophone requirements is sufficient for all other languages', are precisely what they must not make.

To be sure, "the real necessity of any punctuation needs extremely careful examination". Equally certain is that some ASCII punctuation marks are not available for inclusion in domain names for absolutely compelling technical reasons. Such restricted protocol elements are represented using what also happen to be English punctuation marks. This has the obvious further effect of limiting the kinds of punctuation that can appear in domain names. Although the selection of these elements was initially devoid of cultural intent, as with many other aspect of the present matter, conditions have changed considerably in the interim.

Does the unavailability of an apostrophe to indicate English contraction mean that contraction should be prohibited in other languages where it is indicated using different characters that are unique to that purpose? Does the restriction of the range of available punctuation to a single 'separator' in the Latin-based LDH repertoire mean that a similar constraint should be applied to every other script? Should languages using scripts that include letters that resemble ASCII punctuation marks be restricted to subsets of their own alphabets?

In fact, the reference to the Ethiopic wordspace was preceded by careful consideration of factors such as these. Although it does resemble a colon, the two belong to separate scripts that are about as graphically distinct from each other as can be (http://www.unicode.org/charts/PDF/U1200.pdf). Although the hyphen in English and the wordspace in Ethiopian can serve the same function in their respective languages, the hyphen is meaningless in Ethiopian. Making it available would also require breaking the 'one label - one script' restriction. By virtue of the same restriction, the wordspace cannot appear in a Latin string, or anywhere else other than in a sequence of Ethiopic characters. (Nor can a colon appear in an Ethiopic string.)

Telling Ethiopian name holders that they cannot use their wordspace because it might cause confusion in a context in which it can never appear, is a poor way to demonstrate concern for linguistic equality. Seen from the opposite persepctive, if our interest is in preventing user confusion, we need to consider the need for restricting the range of punctuation marks available within the Ethiopic script. All we've been asked for so far is the one -- corresponding in every way to the H in LDH, but in accordance with the needs of one of the many communities to which Latin script is foreign.

Cary Karp
dotMuseum

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

submit APTLD comments to draft IDN Guidelines v.2.0

- *To*: <idn-guidelines@xxxxxxxxx>
- *Subject*: submit APTLD comments to draft IDN Guidelines v.2.0
- *From*: "aclu" <aclu@xxxxxxxxxxxx>
- *Date*: Sun, 23 Oct 2005 10:46:33 +0800

Dear Sir/Madam,

The APTLD has reviewed the "Guidelines for the Implementation of Internationalized Domain Names Draft Version 2.0" which ICANN has published Sept 20, 2005:

http://icann.org/announcements/announcement-20sep05.htm

As an IDN Committee chair of APTLD, I submit the APTLD Comments on "Guidelines for the Implementation of Internationalized Domain Names Draft Version 2.0" to ICANN on behalf of the APTLD.

The APTLD comments has been approved and supported by the APTLD board members. Please reference the attached file to find the contents.

Best regards,
Ai-Chin Lu
IDN Committee chair of APTLD

Attachment: **IDNAPTLD-20051021.doc**
Description: MS-Word document

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index >>>

Comments to draft IDN Guidelines v.2.0

- *To*: <idn-guidelines@xxxxxxxxx>
- *Subject*: Comments to draft IDN Guidelines v.2.0
- *From*: "HARADA Shiro" <harada@xxxxxxxxxxxxxxxxx>
- *Date*: Mon, 24 Oct 2005 08:19:27 +0900

```
Gentlemen,

Congratulations for your efforts towards improvement of the IDN Guidelines.

I would deeply appreciate it if the following principle could also be
included:

Some scripts have characters with almost the same or very similar shape
with certain characters in other scripts.
Digits in Khmer, Thai and Lao scripts are just an example among many.
Even a single script may have varied usages depending on the language
it is used to write.
In the cases of gTLDs and some ccTLDs, multiple scripts may be supported.
When these scripts have such characters with almost the same or very similar
shape,
any decision about one script would affect the others.
Thus such decisions have to be made with full participation of stakeholders
from all the communities of users of relevant scripts and/or languages,
paying special attention to developing countries and valunerable peoples
that have difficulties in active participation by themselves.

Thanks a lot,

HARADA Shiro
Associate Professor
Interfaculty Initiative in Information Studies
The University of Tokyo
```

<<< Chronological Index >>> <<< Thread Index >>>

ICANN ICANN Email List Archives

[idn-guidelines]

<<< Chronological Index >>> <<< Thread Index

Comments on Draft Revised IDN Guidelines

- *To*: <idn-guidelines@xxxxxxxxx>
- *Subject*: Comments on Draft Revised IDN Guidelines
- *From*: Anne-Marie Eklund-Löwinder <Anne-Marie.Eklund-Lowinder@xxxxxx>
- *Date*: Mon, 24 Oct 2005 08:22:42 +0200

```
-----BEGIN PGP SIGNED MESSAGE-----
Hash: SHA1

Dear sirs,

Since you put the deadline late Sunday evening I assume it will be all right to
send our comments to you early Monday morning,
since it is working.

NIC-SE (adminstrating the Swedish top level domain on behalf of the Foundation
for Internet Infrastructure), wishes to express its
appreciation of the initiative and effort behind the current revision of the
IDN guidelines and looks forward to their futher
development. As an IDN registry, NIC-SE is aware of the complexity of the
issues relating to the internationalization of the domain
name space. The guidelines provide a useful basis for the further development
of our own IDN policies and we are pleased to note
that we can implement and support the guidelines as they are currently stated.

There is one extension to the guidelines that we believe requires particular
attention in the short term. The value of the active
area of the IANA Registry for IDN Language Tables would be strongly enhanced if
it clearly differentiated between language tables
contributed by registries that conduct their primary business in the languages
they are describing, and tables contributed by
registries that have collected information about a language's character
requirements from secondary sources. We realize that the
introductory text in each table is usually clear in indicating this
distinction. However, as the registry grows it risks becoming a
new source of confusion to users who may not understand the differences between
two tables nominally describing the same language
but including different arrays of characters.

We also feel that provision needs to be made for script-based character tables,
and for tables that describe language groups. Both
follow directly from the current wording of the guidelines but the IANA
registry has yet to be put to corresponding use.

NIC-SE is preparing two contributions to the IANA registry. We wish to share
our particular understanding of the requirements of
the Swedish language with other registries wishing to support it. We are also
developing a unified table describing all of the
legally recognized minority languages in Sweden. Although this may not be
directly useful to any other registry, it may provide a
basic model for aggregated language tables.

Regards,

Anne-Marie Eklund Löwinder
Manager Security & Strategic Development
```

NIC-SE
Box 5774
114 87 Stockholm
+46 8 4523517
+46 734 315310
http://www.nic.se

-----BEGIN PGP SIGNATURE-----
Version: PGP 8.1

iQA/AwUBQ1x9suJkFdsgpU+DEQKvcACfUC+YXk3Nu7enozug+Rfwx//y7yYAoLG7
xJqFDWrFwEaiXSyMeSwxbBNZ
=IabA
-----END PGP SIGNATURE-----

<<< Chronological Index >>> <<< Thread Index