

UA Measurement WG Meeting

15 December 2022

Attendees

Harsha Wijayawardhana

Jim DeLaHunt

Eng. Moamin Hamad

Sarmad Hussain

Arnt Gulbrandsen

Yin May Oo

Seda Akbulut

Meeting Agenda:

1. Welcome and Roll Call
- 2. What's next with the M4 Action item?
Is IDNA2003 vs IDNA2008 a big deal? or should we look for a more important objective?**

M4: Characterize how much Android platform limits acceptance of IDNs in web browsing (UASG037)

.Identify the UA related constraints on applications running on Android platform)

i.How does UTS #46 differ from IDNA 2008?

ii.Define the outcomes and objectives of the work

3. AOB

Meeting recording

Meeting Notes

Arnt shared the compatibility investigation of IDNA2008's effects on already existing second level domains. If one has implemented IDNA2003 compatible code today, and would like to upgrade to IDNA2008 compatibility, aside from the known and clear advantages, their concerns is what would it break. What would be broken by upgrading to IDNA2008? Arnt had looked into the second level IDN domain names to check their differences with IDNA2003 and IDNA2008

compatible browsers.

There are a few websites with domain names that include unusual characters, and those do not seem to have useful content or transactions. Some of them are domain names for testing purpose and many of them are real, namely “.net” (which has an equal sign and a slash), “.shop” (which has a middle dot), and “.tokyo” (which has a combination of Japanese punctuation with Latin characters), those are illegal in IDNA2008. The chance of broken domains because of IDNA2008 standard is less than (1/10000). To Arnt’s opinion, it would not be harmful to change from IDNA2003 to IDNA2008, changing from IDNA2003 to 2008 offers advantages, and it does not break anything.

Jim thanked Arnt that it is very useful information and requested to share the knowledge to the working group in written form. Jim also said that this particular interest was based on the bug report reply from okHttp, referring to Chrome’s behavior on Android, could be persuasively responded to by Arnt’s findings.

Arnt highlighted that that bug reply seems to say that it wants to be compatible with Chrome. And it is safe to say IDNA2008 would offer advantages and would not break anything that we currently have.

Jim asked about the Eszett character in German language, and about checking on the domain (<http://www.sparkasse-gießen.de>). Arnt explained that DENIC accepted domain names with Eszett characters, and its variants had been checked by DENIC, it might have been taken care of since 2010, so we could conclude that changing to IDNA2008 would not affect these labels.

Jim said in this case, it could be possible to file a response back to the okHttp bug thread and link it to Chrome by sharing what we have researched. We could share with them that there is no downside to changing IDNA2008. Arnt said the trail would lead to the Android source code; what okHttp actually wants to be compatible is Android’s foundation layer, which is adapted to Java version 8 and maintained by Oracle no longer changed, because Oracle is now working on Java version 21. Generally Google does not want to change anything that would break compatibility with Java version 8, however, it might be persuadable.

Jim expressed appreciation to Arnt’s work and helped list out the action items.

Jim suggested preparing an explanation of why the zero-width-joiner (ZWJ) is important for Sinhalese script.

Jim was trying to come up with a few examples of strange-looking-label examples. If W is not existing in the writing system we would use uu.

“Waste World” would have the domain name “Uuaste Uuorld”.
(olemocracy) instead of (democracy) - replaced d by ol

Jim said he will send an email with metaphors explaining why the lack of ZWJ will affect Sinhalese in the perspective of Latin script users.

Jim summarized the action items, which were listed at the end of the report.

Harsha said there are two forms which could be derived from many important words like democracy, truth etc. Jim requested Harsha explaining this would help the rest of the world understand how important this is to have ZWJ. Sarmad said for Arabic script, it is more of ZWNJ.

Seda asked if there is any code or script solution required for changing towards IDNA2008 and if so we may consider having a sample code.

Last meeting it was said **“Our shared code samples should make the implementation of IDNA2008 seem as easy as possible. It would be better to share more IDN and UA compliant codes even if they have bugs, it would always be better to have buggy codes than no codes.”**

Jim said it is not a code problem. We need to show persuasive work and the need. It is more of Android being stuck in the past and Oracle not supporting the past versions of Java anymore. This would be about persuading to make a change out of something frozen. So, we need Arnt’s report and Harsha’s report to back up our persuasive arguments.

Harsha informed the participants that he has shared the affected words to the mailing list “ua-measurement@icann.org”. Seda also requested Harsha to share screen, and Harsha presented his article. Jim thanked Harsha and also requested the list of the names that are broken or prevented by IDNA2003. Jim said it would be helpful to provide the policy of Top Level Domain (TLD) and Second Level Domain (SLD) in the context of Zero-width-joiner (ZWJ). Jim also requested to add the comparison table with three columns, which consists of the U-label, the U-label without ZWJ and the explanation of what had happened to make the paper more concrete. Jim emphasized that showcasing the problems specifically is better than saying things in general when we persuade something.

Harsha sent an attachment to the mailing list: ZWJRenderingEdited2
[Subject “\[UA-Measurement\] Sinhala Yansa and Rakar forms”](#)

Pavanaja also replied to that email with Kannada version.

“Kannada” <> “Canada”.

Harsha added that this script uses ZWNJ and there are more to find out.

Sarmad shared in chat for a complete list of IDN ccTLDs at

(<https://www.icann.org/resources/pages/string-evaluation-completion-2014-02-19-en>) There are two labels for Sinhala script

A-label xn--fzc2c9e2c for U-label ලංකා

A-label xn--xkc2a13hye2a for U-label இலங்கை

Sarmad said in that chat: *for Arabic script, it is more of ZWNJ.*

Sarmad tried to show U-labels اسلامآباد (with ZWNJ) vs اسلامآباد (without ZWNJ)

Jim said this is a very good example because if there is no ZWNJ, it would make it hard for the people to be able to spell it on browsers.

Seda reminded Harsha about the last discussion which were about other browsers Opera, Microsoft Edgy and so on, and Harsha said these browsers do not work. The problems were due to punycode labels.

Jim explained that from a universal acceptance point of view, every piece of software which is IDNA2003 to be upgraded to IDNA2008. There would be many pieces of software limitation, all of them would be eventually pushed on. We would also discover the interlink between software pieces like okHttp linked to Android, so the core Java code would be the leading train to be pushed instead. Now, for the M4 action item, we are able to see the actual problem and the path to persuade, why people should fix it and then fix through the software chain.

Jim also asked a question of persuading tasks that fit this Measurement WG or would it be worth it to pass this to the Technology WG. Jim would prefer having the Measurement WG follow through M4 until persuasive papers are well written, although WG is not sure how to escalate this to business stakeholders. Sarmad said WG can alternatively hand this over to ICANN staff after completing the papers.

Harsha will share the big list. Jim described how the table should look like. Harsha will send it. Either it is a big problem or not. Devanagari has the same problem.

Jim suggested reviewing the FY23 tasks and see if there are other things to focus more than M4.

Sarmad shared in the chat that for Arabic, it is mostly ZWNJ (instead of ZWJ).

Sarmad will share some examples:

اسلام آباد

This is supposed to be one word - Islamabad - capital city of Pakistan.

But if we write it without a space in the middle the two words join.

So ZWNJ could be used here.

Arnt says that he remembers some excellent Arabic examples in Wikipedia.

Being specific is better than general when it comes to persuasion.

Seda recalled adding Opera and Edge to action items for their source code. Jim said we need to push all to switch to IDNA2008.

Next meeting agenda is to follow up on this meeting's action items. And FY23 action plan M2 and M3 <https://uasg.tech/wp-content/uploads/2022/06/UASG-FY23-Action-Plan.pdf>.

SOW for UA curriculum for academia calls for proposals

<https://community.icann.org/display/TUA/UA+Statements+of+Work>

Next meeting: TBD

Action items

No.	Action Item	Owner
1	Send the research and findings about IDNA2003 to the mailing list	Arnt
2	Bug report reply with evidence for okHttp	WG
3	Ask Chrome, Opera, Microsoft Edge to change source code (IDNA2008)	Arnt
4	Write down limitations for Java source code, the future Java and limitation of the Standard	Arnt
5	To share the differences of having and not having the zero with joiner in Sinhala script labels	Harsha
6	Send an email with metaphors explaining why lack of ZWJ will affect Sinhalese in the perspective of Latin script users.	Jim
7	Share some Arabic examples	Arnt / Sarmad