

UA EAI WG Meeting

4 February 2020

Attendees

Harish Chowdhary
Kristina Hakobyan
Mark Datysgeld
Mark Svancarek
Nitin Walia
Pensri Arunwatanamongkol
Sushanta Sinha
T Santhosh
Yao
Pitinan Kooarmornpatana
Sarmad Hussain

Agenda

- 1) Review notes from previous meeting
- 2) Best practices for email admins

Review notes from previous meeting

The WG members reviewed the 21 January 2020 meeting notes. The members were reminded to review [Universal Acceptance Conformance Scoping document](#), the document includes a lot of definitions. The WG agreed to review this document and ensure that both documents are aligned.

Best practices for email admins

The WG continued the [Best Practices for Email Admin](#) document discussion, starting at the variant consideration.

Variant Consideration

It was shared that variant should be identified and blocked for most scripts, except some of the scripts which might need to allow variants as per usability requirements. As of current knowledge, exception cases were Arabic and Chinese scripts. Some examples would be added to the document.

Case Folding

For case folding consideration, the members were informed about existing recommendations from W3C (https://www.w3.org/International/wiki/Case_folding). The WG would review these recommendations for further discussion. It was noted that case folding was irrelevant for domain names as upper case is out of scope by IDNA2008 standards. It was generally agreed that all case folding permutations should be mapped to the same mailbox.

On this topic, the members also added two more points for consideration:

1. Consideration for special characters e.g. Hyphen, Dollar Sign
2. Consideration for digits.

The Unicode script ‘common’ were introduced. As per Unicode Consortium’s [scripts.txt](#) document, special characters that are normally known and may be shared across scripts are defined as ‘Common’. It was suggested that ‘Common’ code points is not considered as ‘mix-script’.

It was agreed to select relevant code points form the ‘common’ script type for further discussion.

“xn--”

The WG discussed and agreed that though the member were not aware of any software converting ascii to Unicode using Punycode algorithm for mailbox names, there has been attempt in the past. Therefore, the mailbox account name beginning with “xn--” should be disallowed.

The topic of normalization form was brought up. Normalization is a Unicode standard defining normalized forms of a glyph as it can be encoded in multiple ways. There were various type of standardized forms for example Normalization Form D (NFD) and Normalization Form C (NFC).

Example: Å can be encoded by [212B](#) or [00C5](#). Both encodings would be normalized to the same NFD and NFC, see graphic below. This is useful for preventing a same visual mailbox account leads to two different mail boxes.

Source	:	NFD	NFC
Å 212B	:	Å ◌ 0041 030A	Å 00C5
Source	:	NFD	NFC
Å 00C5	:	Å ◌ 0041 030A	Å 00C5

Standard software usually already encoded Unicode normalization in their libraries. However, the WG might need to select which form to be recommended for mailbox account names.

The WG noted that a common reference for variant definition should be discussed. A placeholder was added to the document.

Next meetings: Tuesday 11 January 2020 UTC 0500-0600.

Action items

No.	Action Item	Owner
1.	<i>Review following documents for discussion</i> <ul style="list-style-type: none">a. <u>Universal Acceptance Conformance Scoping document</u>b. <u>W3C Case Folding Introduction and Recommendations</u>c. <u>RFC 6530 Overview and Framework for Internationalized Email</u>d. <u>Script definitions by Unicode</u>e. <u>Unicode Normalization Forms</u>	ALL