

UA Tech WG Meeting

11 April 2022

Attendees

Satish Babu
Jim DeLaHunt
Mahesh Kulkarni
Chaitanya Gadam
Dhananjay Garg
Mark Datysgeld
Sanjay
Seda Akbulut

Agenda

1. Welcome and Roll Call
2. UA Community Feedback on the completed Work: [“UA Readiness of 2000 Websites Globally”](#)
3. AOB

Meeting Notes

Seda shared that guest speakers will be joining today’s meeting as well. Satish shared a brief summary of what was being done and handed over the meeting to Mahesh to present a presentation on “Pilot Project for Large-Scale UA Remediation Campaign (EAI Support of Global Websites)”. Mahesh presented the status of testing of websites.

(Feedback from the Tech WG that needs to be incorporated in the final report are highlighted [in blue and*](#).)

Satish asked why ascii@ascii.ascii was not tested this time. The email we had under this category was ending with .com in the 2020 test. Mahesh responded that .com emails are assumed to be working in almost all websites. That’s why it was excluded in the test.

Satish also asked about the reason for lack of comparability between the test sets. The data sets for the email addresses used in the tests are not the same in the

2020 and 2022 tests. Mahesh responded that the tests are provided by the UASG team and Seda can answer it better. Seda confirmed that we used different email addresses. For the specific reason she said she would revert back. *(Follow-up information: Having specific email addresses for this study allows for a way to even see if we get a response back, if needed) In this study, we have more categories in EAI.)*

Mahesh presented that some of the websites, that have unethical and porn content, could not be tested in India due to blockage and censorship. This was something contrary to the requirement in the SOW. Satish commented that in terms of the DNS ecosystem the content does not matter. However, it is well understood that unethical websites are not accessible in India. Satish shared that we should discuss it next time with side steps, such as using VPN. Satish asked comments of the participants regarding addressing this in the next SOW. Mark commented that for the next SOW he would like to see all websites other than the ones serving malware tested. The DNS is agnostic to content. Jim mentioned that in some parts of the world, porn is not regarded as unethical, and we are not in a position to judge people. *(This should be further discussed in the next study.)*

Jim asked Mahesh if he had a spreadsheet with results of deduplication and checking of websites over the past years. According to Jim such spreadsheet gives a longitudinal study of individual URL or domain name. Mahesh responded that all the database is stored. Satish appreciated this thought.

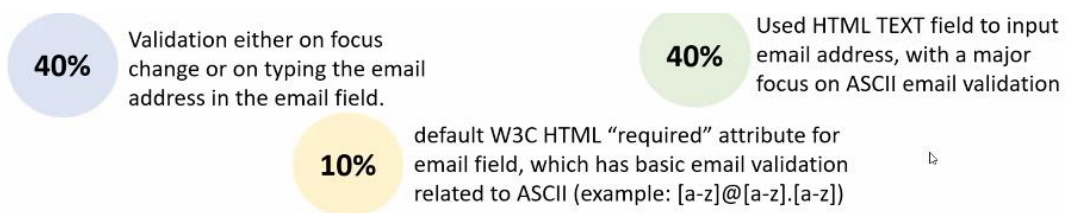
Mahesh explained that they have additional 10 websites just in case there are some sites that cannot be tested. All websites are given a unique ID with P code for 2019 websites data, N for 2020 website data, A for additional Alexa top 380 websites which were included after deducting duplicate, unethical or untestable websites. A total 550 websites were subjected to basic website “Functionality” and “Test email field Identification”.

Mahesh presented the observations related to Re-Testing of Websites like some websites requiring valid mobile number, some are country specific sites which require national identity number, and some sites did not show error message, usually the page was redirecting to the home page. In that case it was not possible to understand whether the email was submitted successfully or not.

Regarding the focus change, when there is no success or error message, usually the page was redirected to the home page. In that case it was not possible to understand whether the email was submitted successfully or not.

Mahesh shared that complete registration was done for C1 and C2 category. Some websites needed some other kind of verification and implemented that. Some websites were tracking the activity and not allowing multi registrations. Some required credit card details and registration was not allowed.

Mahesh then presented the observations related to functionality and email testability. Few websites have a "Login/Registration" page view on dialog/popup. Multiple clicks and views are required to reach the Registration Page for locating the email field. However, the URL remains the same, and hence it is necessary to note all the steps for carrying out re-testing. Few of the websites had "email-based login and registration", however, it seems that they have now changed login to Mobile number plus OTP combination. Also, for new users, registration with email is disallowed, however existing users continue to use mail/UserID/mobile for login. Few websites have upgraded to "Mobile plus OTP" based login instead of email address. Also, the email field is not observed in the contact /support/feedback page. Hence these are marked as non-testable. Testing team not conversant with the language of the websites faced challenges during navigation and email field identification. Combination of machine translation plugins such as 'Google Translate' and "Google lens" Android application were used for translation. Google translate as well as Google lens did have their own issues.



*This chart depicts how the validation works in websites. It shows what mechanisms generally are used to validate whether they are the right mechanism or not. These percentages do not add up to 100%; they all are independent.

*The first 40% is about JavaScript, the 2nd 40% is HTML TEXT which takes email ID as text input, and the 3rd 10% uses input type equal to email ID but validating ASCII. Mahesh also shared that this division is how it works and not the result of testing. Satish mentioned that we are talking to W3C to correct that. Mark D. supported the idea of engaging with WHATWG and W3C about HTML5.

Mark D. and Jim asked about what the 10% category exactly is, which refers to HTML email input type for basic email validation related to ASCII. Mahesh explained that W3C made this requirement mandatory by default. Satish summarized that this 10% shows us the `<input type= "email">` and validation is done by the browser itself, and not by the programmer. Mahesh added that this attribute is required.

Jim asked Mahesh if it is the case that websites in this category are using the html input elements with the `<input type= "email">` attribute. Mahesh responded that 10% of the people using email as an attribute made it mandatory.

*Jim also asked if the browser behavior is different if input type is equal to email id. Mahesh said the browser doesn't view changes because of that, and he doesn't really know why they put the required attributes. Jim explained that the reason for making it mandatory is probably because the developer wants to make sure the user inputs an email address. There is a bug report UASG opened with W3C about input element: "Validating internationalized mail addresses in `<input type="email">`" #4562, <https://github.com/whatwg/html/issues/4562>, which limits the email address to ASCII.

*Mark D. mentioned that this would be a use:

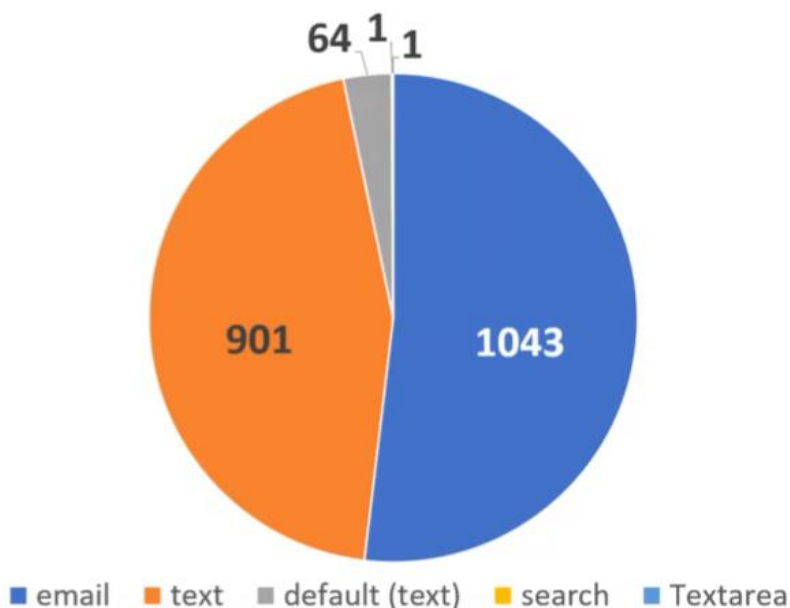
```
<input type="text" id="username" name="username" required>
```

In the above markup, HTML5 would automatically send a flag which you can manipulate is CSS to mark the field as red or whatever. But the behavior itself doesn't change. *He suggested that maybe the team wants to point out that these fields are actually blocking people due to ASCII barriers. The problem would be that there are other methods to require via JS and PHP for example.

Mark D. mentioned that our previous finding stands that in the more basic formats, we saw a jump due to HTML5 adoption in 2019-2020. It became stuck since then because that standard wasn't updated and we haven't managed to start changing code bases to be UA-Ready yet. It shows just how important UA-Tech WG is, and how much we still have to accomplish. Satish agreed to that.

Mahesh presented the EAI acceptance rates comparison for different years and for different series. It was observed that there is no major difference between the EAI acceptance rates of the 2020 and 2022 reports. However, the 2022 EAI acceptance rates are marginally lower than in the 2020 report (UASG027).

It is interesting to note that for the email category “ascii@idn.ascii” there is an increasing trend in acceptance rate of 2019 vs 2022 and 2020 vs 2022. Certain websites have put a limit on the maximum number of characters the user can enter in the field meant for email ID. Mahesh mentioned that for the [unicode@ascii.ascii](#) type of address, the characters are longer than the previous test. Therefore there is a decrease in the EAI Acceptance.



It is observed that few of the websites have whitelisted specific TLDs (i.e., only emails from those domains are permissible). While inspecting the email field in various forms of the websites, it was observed that the email field had values for the HTML <input> element “Type” attribute.

Satish asked Mahesh about the choice of [a-label.a-label](#) not eligible for testing purposes. In practice, people use them as IDN, not in punycode. Mahesh shared that he is not clear about it but maybe it is to check if punycode is accepted or not. Satish agreed to that sharing that it might create a problem for acceptance because 71 characters (including the extension) is too long.

*Jim mentioned that it would be interesting to come up with a best practice recommendation for A-label length to support. Having a max of 25 characters is

too short. He suggested that maybe we should test A-labels of maximum label length supported by DNS, or whatever the maximum label length of all UASG004 domain names in A-label form. Mark D. Agreed to that. Satish mentioned that we have not tested the length aspect so far. And perhaps we should test it in the next round of tests.

Mahesh also shared that the next step is to reach out to the websites that are not UA ready. Satish appreciated the efforts of Mahesh and his team. He asked him to share the slides with everyone.

Next meeting: Monday 25 April 2022 UTC 1600-1700

Action items

No.	Action Item	Owner
1	Sharing Evaris report and presentation with the Tech WG	Mahesh / Seda