

Internationalized Domain Names (IDNs) : A Key to Inclusive and Multilingual Internet

Background:

- Until recently, the Root was limited to a set of characters conforming to US-ASCII (American Standard Code for Information Interchange) or "Latin" alphabets. This changed with the introduction of Internationalized Domain Names (IDNs), which introduced top-level domains (TLDs) in different scripts and enabled Internet users to access domain names in their own language. The approval of the IDN country code Top-Level Domain (ccTLD) Fast Track Process by the ICANN Board at its annual meeting in Seoul, South Korea in October 2009, enabled countries and territories to submit requests to ICANN for IDN ccTLDs representing their respective country or territory names in scripts other than US-ASCII characters.
- Internationalized Domain Names (IDNs) were first launched in 2004 called IDNA, and now based on IETF standard RFC5890 published in 2010. It uses Punycode encoding algorithm to represent non-ASCII characters found in Arabic, Chinese, Cyrillic, Hindi and other languages, into ASCII character domain names that the DNS system can resolve. This allows Internet users to type a domain name in their local script using their native language, instead of an English version [2].
- For many years, the simple task of typing a web address into the browser isn't possible for people who don't know English but introduction and implementation of IDN's have reinforced the fact that Internet is not only English driven other languages can play important role to connect the next billion non-native English speaking population as 83% of the world's population is estimated to be non-English speaking, yet for years they have been unable to communicate in their own languages via Web and e-mail addresses [1].
- Therefore one of the most important innovations for the Internet since its beginning will be the introduction of top level Internationalized Domain Names (IDN TLDs). These IDN TLDs will offer many new opportunities and benefits for Internet users around the world by allowing them to establish and use domains in their native languages and scripts. This is enabled by allowing domain names to have characters from different scripts, beyond the letters (a to z), digits (0 to 9) and hyphen (-), as encoded by the Unicode standard and as allowed by relevant IDN protocols [3].

- The new generic Top-Level Domain (gTLD) Program, approved at the ICANN meeting in Singapore in June 2011, will allow for the first time the addition of IDN gTLDs into the root zone. Thanks to the policy of opening up scripts other than Latin by ICANN, a flood-gate of new languages and scripts has opened up and domain-names will become truly multi-lingual in nature. Benefiting from this new policy, India has taken up the challenge of providing IDN's in Indian scripts and languages for the 22 official languages of India.
- The formulation of a policy document for India to provide Internationalized Domain Names in the 22 official languages has been nearly 5 years in the making. Started in 2005, the policy has been elaborated over the years to ensure that the eventual users will have as safe as an environment as possible when they register their names in an Indian language using their native script. Seven Indian languages (Hindi, Tamil, Telugu, Gujarati, Bangla, Urdu and Punjabi) have already been proposed to ICANN and IANA and the ccTLD for the country name "India" in these languages have already been approved and delegated into the DNS root zone.
- Since scripts do not share the same composition rules and have their own "grammar of composition"; it was in the fitness of things, that ICANN felt that the creation of "test cases" in six scripts would allow for a better perception of the problems as well as issues involved. The scripts chosen for study (apart from Latin): Greek, Cyrillic, Arabic, Devanāgarī, Chinese reflect in fact the 4 major writing systems of the world Abugidas (Greek and Cyrillic), Abjads (Arabic), Akshar or Alphasyllabaries (Devanāgarī) and Phonetic-Semantic (Chinese).

How IDNs Work?

- The Internet was not designed to be multilingual. The domain name system (DNS) was intended only to support the characters a through z, A through Z, 0 through 9, and the hyphen. Upgrading the DNS to support upwards of 100,000 characters is no trivial feat. In fact, the DNS has not been upgraded. The name servers scattered around the world—the servers that translate domain names into their numeric IP addresses—still only support ASCII characters. But what has changed is the addition of a sort of multilingual translation service at the front end (the user's browser) to supplement the DNS.

- When an IDN is input, the web browser algorithmically converts it into an odd-looking combination of ASCII characters (known as punycode). For example, the IDN for McDonald's Russia (макдональдс.рф) is converted to the puny code equivalent <http://xn--80aalb1aici8a5i.xn--p1ai/>. It is this puny code string that is sent to the name server. When an IDN is initially registered, it is this puny code address that is stored within the name server.

Punycode in My Browser:

- Users of IDNs may notice puny code showing up their web browsers, such as this example: xn--h2brj9c puny code for भारत. This is usually intentional. With Internet Explorer, Chrome, and Firefox, the IDN is generally displayed in punycode format unless the language preference of the browser matches the language of the IDN.
- There are real and perceived security fears that IDNs may be used to trick users into visiting phony websites. For example, paypal.com and paypal.com are actually two different domains (the first uses the Cyrillic character "p"). **To minimize this risk, some IDNs, such as the Russian IDN, do not permit mixed-script domain names. Of course, it's important that registries enforce these limitations, which remains to be seen.** For the time being, expect inconsistencies in how different browsers handle different IDNs

Obstacles to IDN Adoption

- While IDNs are in many ways a natural evolution of the Internet, most Internet users are completely unaware that they are even a possibility. And companies have largely been slow to promote them. In some cases, executives are concerned that promoting an IDN address may conflict with existing efforts to promote the Latin-based addresses. The thinking goes: *Even if the current URLs are not ideal, why confuse users any more than we have to?*
- In addition, the emergence of social networking platforms also may hinder the growth of IDNs. For example, Twitter doesn't currently recognize IDNs (though Facebook does).
- There are also a number of technical and logistical challenges still to overcome. For example, there are concerns over how best to manage domains that may be represented by "variant" characters. For example, China has two distinct IDNs.

- One IDN is in simplified Chinese script while the other is in traditional script. The question that must be resolved still is whether or not a domain registered with one IDN should be aligned with the other. This is just one example of numerous variant issues that must be resolved by ICANN in the coming months (it has working groups studying the challenges right now).

Challenges to Supporting IDNs on Websites

Although IDNs are a long way from becoming commonplace, web developers should begin to consider the implications of supporting IDNs. Looking ahead, one can no longer assume that a URL will support only ASCII characters. Here are some challenges to keep in mind:

- In a world of multilingual URLs, web developers will need to balance security with usability. URL input fields that automatically block non-Latin characters may need to be modified. How should IDNs be promoted? For instance, McDonald's in Russia still displays its Latin URL on its home page instead of the Cyrillic equivalent. Should companies display both URLs, and which URL should be the “front door?” There are no easy answers here, but the questions need to at least be asked.
- If the domain is in a different script, should subdirectories also be in the same script? Many times we have seen the top-level and second-level domains are in local languages but the subdirectories are in Latin characters. If a company is going to ask users to input long URLs to go directly to subdirectory sites, usability needs to be taken into account.

Criteria:

- **Security:** Registries should be resistant to hacking, and must also withstand denial-of-service (DoS) attacks. The hardware and software requirements for such secure systems are high.
- **Reliability:** Various user entities – government, businesses, citizens– rely on the Web sites and e-mail messaging that domain names enable. A Registry should provide reliable service, with 100% uptime.
- **Scalability:** The Registry infrastructure should accommodate growth without interruptions.
- **Adherence to** - The Internet functions well only when the concerned Technical Standards Agencies collaborate to create and maintain interoperable technical

standards. Registry should adhere to such standards and practices announced by ICANN and other relevant technical standards bodies.

Issues with IDNs

- No ubiquity
- No consistent user experience
- Lack of awareness
- Lack of trust

With the emergence of IDNs, we are inching closer to a more linguistically local internet in which users no longer have to leave their native languages to get where they want to go. And though there are many obstacles ahead and there is a positive steps towards making the internet through making the internet truly accessible to the world.

Expected Outcome of the Workshop:

- Sharing of information regarding the latest developments and challenges in creation of multilingual internet ecosystem in the country and update on efforts being put in by various key players of the ecosystem.
- Providing a platform for the industry specific concerns/ challenges and on how Government can proactively participate and support the industry to develop business models to sustain multilingual internet ecosystem in the country.
- The industry is expected to be informed of the bigger picture and come out with products and services in local languages.
- Providing a platform for all stakeholders in the ecosystem to facilitate technology and information transfer.

References:

[1] <http://www.afilias.info/idnemail>

[2] <http://www.afilias.info/IDN-Backgrounder>

[3] <https://www.icann.org/resources/pages/idn-2012-02-25-en>

By: Reshmi Hariharan & Harish Chowdhary

