Internationalized Domain Names in Applications (IDNA) and Universal Acceptance (UA)

Harsha Wijayawardhana B.S. in Biochemistry (Miami), CITP (UK), FBCS (UK)

Current Chair, Local Language Working Group (LLWG) of the ICTA

1. Introduction

Internet Corporation for Assigned Names and Numbers, which is known worldwide as ICANN, concluded its Annual General Meeting, ICANN 75, in Kuala Lumpur, Malaysia, on the 22nd of September, 2022. The multilingual Internet became a central topic at the conference. Presently, 5.8 million Internet users are in the world; it is a percentage of more than seventy percent of the world population. ICANN plans to attract another 1.8 billion users to the Internet by popularizing Internationalized Domain Names and adopting Universal Acceptance. Presently, it is estimated that there are more than 4.7 billion social media users spread across the world.

Until the mid-nineties, Latin Scripts alone dominated the Internet, and by then, English had become the widely used language on the Internet. With the development of Unicode technology, Chinese, Japanese, Korean and Arabic users clamored for content in their languages. However, English remains the preferred language on the Internet to the present day, and English content occupies more than 58% of the total content on the Internet. Internationalized Domain Names have also become a hot topic due to the rise of non-Latin Script users on the Internet. Chinese, Japanese and Korean (CJK) content began filling the Internet space at the turn of the second millennium. Slowly and steadily, CJK content began to grow, and currently, the number of users who work in Chinese has reached the same number as in English.

2. Internationalized Domain Names (IDNs)

An Internationalized Domain Name is a domain name that contains a label displayed in the whole or part of, in Non-Latin Script or Alphabet such as Bengali, Sinhala, Chinese, Korean, etc. IDNs were mooted and proposed by Martin Durst in

1987 and were implemented by Tan Juay Kwang and Leong Kok Yong under the guidance of Tan Tin Wee in 1990. Domain Name System is a lookup table for translating a Domain Name on the Internet to an IP Address and restricts Domain Name entries to ASCII. Non-Latin Script Domain Names, which are in two-byte code- Unicode, are entered into the DNS lookup table in Punycode. Punycode was developed to transcode two-Byte Unicode to a limited subset of ASCII consisting of letters, digits, and hyphens, which is known as the Letter-Digit-Hyphen (LDH) subset. තීක්ෂණ.ලංකා is the corresponding Domain Name, Theekshana. lk, in Sinhala Script in IDNs ccTLD, and තීක්ෂණ.ලංකා transcodes into the following Punycode: xn--3zc8ae9ftb0c.xn--fzc2c9e2c.

ICANN has appointed a series of committees since 1990 to make IDNs technically viable. In 2003, ICANN released Internationalized Domain Names in Application (IDNA) or IDNA 2003, which was a mechanism defined for handling non-ASCII domain names. In July 2003, ICANN issued the guidelines for the use of IDNs, and .jp began registering IDNs using the above system. In 2008, ICANN released IDNA2008, an enhanced version of IDNA 2003. With the full-blown development of Arabic support in Unicode, ICANN laid the groundwork for the release of Arabic Top Level Domain (TLD) in Arabic Script. Indians began working on their twenty-two official scripts. In 2009, ICANN approved the first eleven ccTLD IDNs. With the first batch of these, ICANN approved releasing ccTLD in Sinhala .ලංකා (.lanka in Sinhala) and Tamil equivalent .இலங்கை.

Figure 1

3. Generation Panels

ICANN's decision to introduce generic Top Level Domains (gTLDs) in non-ASCII scripts created the need to formulate rules to identify variants in the same Script and Scripts with similar-looking characters or glyphs. To fulfill the above requirement, ICANN formed volunteer community groups for seventeen scripts, which were known as Generation Panels, to develop and maintain Label Generation Rules for the Root Zone in respect of IDNA labels. These members, who were responsible for respective scripts, developed rules that define, for a given script, which code points are acceptable for the root, which labels were variants of each other—and where variants exist—whether such variant labels could be delegated. The same committees formulated rules for the second level subsequently.

Sinhala Generation Panel was formed in 2018, and the committee submitted the rules to the Integration committee in 2019. Sinhala Generation panel completed its task and committed its rules to Root Zone LGR by the end of 2020. ICANN announced the integration of 26 scripts to the RZ -LGR successfully at the ICANN 75 on the 21st of September, 2022. Currently, ICANN has released version 5 of LZ-LGR.

4. Universal Acceptance

ICANN 75 dominated the Universal Acceptance (UA) and allocated many sessions on it. Ram Mohan coined UA in 2001, and it is defined as every Top Level Domain (TLD) must function within applications irrespective of the Script, Device, or how new the TLD is. In other words, UA ensures that all Domain Names, including new TLDs, Internationalized Domain Names (IDNs), and Email Addresses are treated equally and can be used by all Internet-enabled Applications, devices, and systems. ICANN expects that the successful implementation of UA will bring more than one billion new Internet users.

Recently, ICANN accepted the proposal submitted by Theekshana on the implementation of UA in Sri Lanka. A subcommittee of the Local Language Working Group (LLWG) will act as the Local Universal Acceptance Steering Group (Local UASG).

5. Conclusion

Sri Lanka has benefitted immensely from the use of Local Languages on the Internet. The implementation of UA will attract more Sri Lankan users to the Internet, and Sri Lanka has to start using two local ccTLDs, dotLanka in Sinhala Script and dotIlangai in Tamil Script. Hopefully, the kick-off of the UA program in Sri Lanka will make most applications UA-ready, including all email servers and other applications developed in Sri Lanka. Hopefully, since Sinhala is already integrated into RZ-LGR when the applications are called for the next round of IDN-gTLDs, Sri Lanka can apply for gTLDs in Sinhala and Tamil scripts.