

Embracing the Next Billion Users of the Internet

Ms. Pitinan Kooarmornpatana
ICANN IDN Programs Manager
pitinan.koo@icann.org

Challenges for the next billion users?

The next billion Internet users are likely to be non-English speakers.

Even though website content can now be in a local, non-English language and script, domain names have traditionally consisted of letters a-to-z, digits 0-to-9 and hyphen. This can become an obstacle for speakers of different languages and scripts to access the content.

The Internationalized Domain Names in Applications (IDNA) 2003 standard was devised by the Internet Engineering Task Force (revised to [IDNA2008](#)) to enable domain names in different languages and scripts that are supported by the [Unicode](#) standard. Domain name consists of labels at multiple levels separated by dots. Though the IDNA standard allowed for labels to be represented in multiple scripts, it was only in 2009 that the top-level domain (TLDs) labels, such as “com”, “org”, “in”, “sg”, etc., were allowed for delegation in local languages and scripts for country code TLDs (ccTLDs). In 2013, this was allowed for generic TLDs (gTLDs).

During the applications of IDN ccTLDs and IDN gTLDs, the Internet community identified cases in which different labels may be considered “same” or indistinguishable by the end-users. For example, labels in Chinese can be written in two different ways – using the Traditional Chinese or the Simplified Chinese characters. This is called variant forms. Similarly, some words in Latin script can be written in exactly the same visual form using the Cyrillic script. Unless such variant labels are identified and managed, they can pose usability and security challenges subsequently, e.g., pose a significant phishing potential.

⊙ Example of within-script variant labels (Arabic script)

شبكة (06C3 06A9 0628 0634)

شبكة (0629 06A9 0628 0634)

شبكة (0629 0643 0628 0634)

⊙ Example of within-script variant labels (Simplified Chinese and Traditional Chinese)

名称 (540D 79F0)

名稱 (540D 7A31)

⊙ Example of cross-script variant label (Latin script and Cyrillic script)

epic (0065 0070 0069 0063)

epic (0435 0440 0456 0441)

⊙ Example of cross-script variant label (Telugu script and Kannada script)

గళము (0C17 0C33 0C2E 0C41)

ಗಳಮು (0C97 0CB3 0CAE 0CC1)

Figure1: Sample of variant labels, the visual forms and their Unicode code points.

Solution for enabling multilingual domain names

In 2010, the Internet Corporation for Assigned Names and Numbers (ICANN) community identified that in the Domain Name System (DNS) environment, there is no accepted definition for what may constitute a variant relationship between top-level labels. Subsequently, a Root Zone Label Generation Rules (RZ-LGR) Procedure¹ was developed to determine valid domain names and their variant labels for a particular script. Based on this Procedure, LGR for each script is defined by the relevant script-community panel called a Generation Panel (GP).

Currently there are 28 scripts to be supported at the top-level of the DNS, of which 15 scripts have completed their RZ-LGR proposals² for further review and integration. Many more GPs are working towards finalizing their work³.

The solution for a particular script consists of three analyses. It should contain:

- The characters valid for use in the DNS,
- The variant rules for each character and
- The additional constraints on the entire label which are crucial for defining well-formed labels for complex scripts and managing the number of usable variant labels.

Each GP needs to carefully evaluate the possibilities and propose a conservative solution to address the usability and security considerations.

Neo-Brahmi Generation Panel

In this context, in 2015, the Neo-Brahmi Generation Panel (NBGP) was formed to develop such rules for nine scripts used in South Asia, including Bangla, Devanagari, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Tamil and Telugu. The NBGP members comprise of more than 60 experts in technology and linguistics from Bangladesh, India, Nepal, Singapore and Sri Lanka. The NBGP has already finalized the proposals for many of these scripts, which are currently undergoing public review (see www.icann.org/idn) and aims to finish its work in the coming months.

The envisioned future

With the support of different communities across the globe, ICANN helps to define and document the rules to determine valid domain name labels for the top-level and identify their variant labels. This allows complete domain names, including top-level domains, to be available to communities in their own languages and scripts in a secure and stable manner, promoting better accessibility of domain names across the world.

References

- 1 <https://www.icann.org/en/system/files/files/lgr-procedure-20mar13-en.pdf>
- 2 <https://www.icann.org/resources/pages/lgr-proposals-2015-12-01-en>
- 3 <https://www.icann.org/resources/pages/generation-panel-2015-06-21-en>



About the author: Pitinan Kooarmornpatana is a Computer Engineer professional. Pitinan joined ICANN in August 2017 as an Internationalized Domain Name (IDN) Programs Manager. She is currently focused on facilitating various Script Generation Panels to develop the Label Generation Rules proposals for the root zone.

ICT as applied to social protection

Information and Communications Technology (ICT), as applied to social protection, falls into five broad categories:

- **Technical or incremental.** Automating or replacing manual routines and paperwork with digital technologies, such as the use of stand-alone computers.
- **Sustained.** Bringing about long-term organizational and administrative improvements in efficiency, accuracy, and targeting, such as the use of databases and analytical software programs.
- **Disruptive.** Fundamental restructuring of the way that programs are organized, often involving an internal shift of ownership of projects, which could, for instance, result in the merging of government agencies. The integration of databases may be part of this process.
- **Radical.** Providing greater stakeholder influence through web-based information systems, and the involvement of nongovernment organizations or community associations.
- **Transformative.** Facilitating a wholesale change in the approach toward social protection, such as web-based systems of self-declaration and assessment. Modern taxation systems often move in this direction.