

# Breaking the internet language barrier

One of the ways ICANN is trying to make the use of internet and its content accessible is through the introduction of Internationalised Domain Names (IDN)

**V SRIDHAR & AJAY DATA**

Sridhar is a professor at IIT, Bangalore, and Data is chair of UASG and co-chair of ICANN Neo Brahmi Generation Panel

**DOMAIN NAMES PROVIDE** the important link between users and content on the internet. When we type "meity.gov.in" to access the web site of the ministry of electronics and information technology, we present the domain names interspersed by "dots" to the Domain Name System (DNS) that resolves them into machine readable Internet Protocol (IP) addresses for access. However, the label to the right of the dot (also called as the Top Level Domain, TLD) in any domain name is of utmost importance as this is administered and governed by the multi-stakeholder community model of internet governance under the aegis of the Internet Corporation for Assigned Names and Numbers (ICANN). The TLDs are registered in the "root zone" of the internet with corresponding Label Generation Rules (LGRs) for the stable functioning of the domain name system.

The DNS started with 6 TLDs in the 1980s that consisted of just 3 Latin characters; subsequently, country code TLDs were made available (such as 'in' for India). Subsequently, generic TLDs that can have longer labels such as 'cookingchannel' and 'travelersinsurance' with certain restrictions were approved under the new gTLD programme of ICANN in 2012. While these developments were taking place, software and application developers, network engineers, and domain name registrars had to re-engineer their existing programmes to recognise the shift from legacy 2-3 character TLDs to lengthy TLDs. This, referred to as "Universal Acceptance (UA)" principle, enables any TLD, once defined in the Root Zone, to function within all applications regardless of script, number of characters, or how new it is.

Meanwhile, the internet penetration across countries started growing exponentially. However, despite the growth of

internet in non-English speaking countries, the content on the internet is still predominantly available in English followed by Chinese. One of the ways ICANN is trying to make the use of internet and its content accessible, especially among non-English speaking internet users, is through the introduction of Internationalised Domain Names (IDN). The solution was in adopting Unicode standard that provides a unique number for every character, no matter what platform, device, application or language. Realising the importance of Indian language specific TLD, the government of India obtained (.Bharat) ccTLD in Devanagari script in 2011, which was introduced to the public for domain registration in August 2014. Variations of Bharat TLD is now available in 15 scripts including Bengali, Tamil, Telugu, Gujarati, Urdu and Gurmukhi.

As per the IDN World Report 2018, where IDNs are in use, the language of web content is more diverse than it is with traditional ASCII domains. IDNs help to enhance the linguistic diversity in cyberspace and seem to be accurate predictors of the language of the web content. The report also points out that Han (associated with Chinese language), Latin, and Cyrillic scripts represent nearly 90% of all registered IDNs. Major world scripts such as Arabic and Devanagari, which support some of the world's top 10 most spoken languages, are yet to be substantially represented in IDN. To make Indian language specific TLDs possible, the Neo-Brahmi Script Generation Panel (NBGP) was formed by nine communities in 2015. NBGP is developing Root Zone LGR for Bengali, Devanagari, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Tamil and Telugu scripts. Once implemented, domain names in the above Indian languages can be registered to address the non-English inter-

net users in the country.

Apart from accessing web content, domain names are also used for email addresses and a host of other Internet applications. Hence, UA requires that software applications should be updated to accept the new gTLDs and IDNs. Once implemented in full, end users can use applications with the new domain names without compromising on functionality and performance. In a recent study by Analysis Mason, it is estimated that UA would provide an economic benefit of close to \$10 billion. To promote UA, ICANN has formed the Universal Acceptance Steering Group (UASG) which is spreading awareness of the ramifications of new gTLDs and IDNs amongst all stakeholders. Companies such as Google, Microsoft, Xgenplus have started supporting email address internationalisation (i.e. email address for IDN domain), thus providing UA-ready messaging services. Hence it is very important to educate software developers, engineers, domain name registrars and registries on the importance of UA in the context of new gTLDs and IDNs.

Currently India has more than 50% of the 900+ TV channels that broadcast in regional languages; Hindi language newspapers are the largest in terms of readership. It is time to prop up internet content with Indian languages. The internet 1.0 was the one without Web; the 2.0 variant with the hyperlinked web provided the much-needed network connectivity of content; 3.0 enabled access to the web through mobile and portable devices; internet 4.0 breaks the language barrier for both content and access. This revolution provides a fertile ground for development of content and applications based on the economic, social, cultural and linguistic diversity of the internet population around the globe.

