Label generation rules for the Internet’s root zone for African scripts

Entebbe - Uganda | 03 -04 November 2016

Yaovi Atohoun, Stakeholder Engagement & Operations Manager - Africa
What is ICANN?

The Internet Corporation for Assigned Names and Numbers (ICANN) is a global multistakeholder, private sector-led organization that manages Internet resources for the public benefit.

- ICANN coordinates the top-level of the Internet's system of unique identifiers via global, multistakeholder, bottom-up consensus policy processes, with the outcome of those processes implemented via the IANA Functions.
ASCII Domain Name Label

Forming ASCII Labels
Use LDH
- Letters [a-z]
- Digits [0-9]
- Hyphen (LDH)
Label length = 63
Other constraints (e.g. on hyphen)

Forming ASCII Labels
Use only Letters
- Letters [a-z]
Label length = 63
Internationalized Domain Name (IDN) Labels

Syntax of IDN Labels
Valid U-Label: Unicode code points as constrained by IDNA2008
Valid A-Label - “xn--” followed by punycode of U-Label of length 59

Syntax of IDN Labels
Valid U-Label, further constrained by the “letter” principle for TLDs
Valid A-Label
IDN Country Code Top-Level Domains

ccTLDs: 57 for 39*

* Successfully evaluated IDN ccTLDs for total countries and territories
Label Generation Rules Process

- Generation Panel 1
- Generation Panel 2
- Generation Panel 3

Integration Panel

- One Generation Panel per writing system (script)
- Needs more work

Rules
- Panel 1
- Panel 2
- Etcetera

http://....TLD

Etcetera
Label Generation Rulesets (LGRs) used to generate domain name labels, per RFC 7940

LGR Toolset allows for the following:
- Create a LGR
- Use a LGR to validate label and variants
- Manage LGRs

Online beta deployment
- Visit https://lgrtool.icann.org/

Open source package(s) release
- Released at github: lgr-core, lgr-django, munidata

User guide available for further details
Status of Root Zone LGR Development

Label Generation Rules (LGR)

Generation Panel Status

- Finalizing
- Seated
- Active
- Forming

28+ Scripts
19+ Generation Panels

Other:
- Hebrew
- Sinhala
- Thaana

Maximal Starting Repertoire (MSR)
Linguistic Diversity in Africa

- More than 3000 languages in Africa
- Writing systems include:
  - Ge'ez (Ethiopic)
  - N’ko
  - Tifinagh
  - Vai
  - …
  - Arabic (introduced)
  - Latin (introduced)
Case Study:
Ethiopic Generation Panel
Development Process by the Ethiopic LGR

1. Submission of a proposal to ICANN with expression of interest to establish a Generation Panel (GP) for the Ethiopic script
2. Exploration of the main document that explains the various pertinent issues with regards to IDN and LGR proposal development procedures and processes entitled “Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR”
The Generation Panel is composed of members from different disciplines with complementary expertise. Members professional background include- technical community who have a thorough knowledge and skills in DNS, its internal makings and operations, Linguists, academia, Information Technology and Computer Science professionals working in the industry, Legal researchers, Cyber law experts, and researchers on natural language processing.

More at https://community.icann.org/display/croscomlgrprocedure/Ethiopic+Script+GP
Activities for the Ethiopic GP

- Defining the code point principles
- Conducting corpus analyses for Amharic and Tigrigna languages using contents published online
- Defining the frequency cut of a character or code point for the corpus data analysed
- Building the code point repertoire
- Defining the code point variant principles
- Establishing the code point variant sets
- Analyses of allocatable vs. blocked variants
- Producing a documentation for the process
- Creating an XML LGR for Ethiopic Script LGR proposal
- Submitting the Final LGR proposal for Ethiopic Script along with its XML specification
1. Identifying languages meeting criteria (e.g. EGIDS scale of 4 and less)

2. Investigating selected languages with respect to IDN requirements per the guidelines
   i. Using rigorous criteria for code points inclusion
   ii. Identifying in-script variants
   iii. Identifying cross-script variants

3. Gathering necessary evidence for proving a character’s contemporary use in the writing system

4. Mapping the textual LGR proposal into machine readable (XML) artifact
Accomplishment by the Ethiopic panel

- establishment in January 2016 after getting approval from the Integration Panel by meeting the requisite requirements.
- three face-to-face meetings in January 2016, March 2016 and May 2016.
- Two days training workshop on the 22\textsuperscript{nd} and 23\textsuperscript{rd} of July 2016. In the training workshop besides panel members several invited participants drawn from academia, ISP and media have taken part.
Draft Proposal for the Ethiopic Script LGR for the Root Zone has been submitted. We are expecting to receive the IP’s feedback in the immediate future. Afterwards it will be a case of mapping the LGR document into machine readable thing- XML using the online tool available. We are planning to conclude the project in two months time.
Conclusion

- Ethiopian Script panel is a case of the role academia can play in Africa
- More people needed in the Latin script panel as many African languages are using this script
- Expect to see more involvement from the Academia.
How to get involved?

Volunteer for your script Generation Panel (GP)
To contribute expertise, contribute to the GP for your script. You can get involved by simply emailing your CV and a brief statement of interest to idntlds@icann.org

Review work through public comments
Sign up for the IDN mailing list vip@icann.org (to sign up, visit https://mm.icann.org/listinfo/vip) and participated in the review of IDN work being done at ICANN through the public comments

Keep yourself updated
Attend regular IDN Program Update sessions at ICANN meetings and sign up on the IDN mailing list vip@icann.org to get updates on the IDN Program at ICANN

Visit icann.org/idn
Engage with ICANN

Thank You and Questions
Reach us at: IDNProgram@icann.org
Website: icann.org/idn

twitter.com/icann
gplus.to/icann
facebook.com/icannorg
weibo.com/ICANNorg
linkedin.com/company/icann
flickr.com/photos/icann
youtube.com/user/icannnews
slideshare.net/icannpresentations