IDN Variant Issues Project Greek Case Study Team Update

Panagiotis Papaspiliopoulos

Greek Case Study Team Members

Vaggelis Segredakis (Coordinator)

Asimina Giannopoulou

Catherine Tsapikidou

Evangelos Melagrakis

Fotia Panayiotou

George Papapavlou

Giannis Papaioannou

Panagiotis Papaspiliopoulos

Current Status

- The Greek Case Study Team developed a Report: "Study of the issues present in the registration of IDN TLDs in GREEK characters"

- The report was posted for comment in the Public Forum on 07 October 2011. See: http://www.icann.org/en/public-comment/idn-vip-greek-07oct11-en.htm.

- Comments will be accepted in the Forum until 14 November 2011.

Structure of the Report

- Introduction-Disclaimer
- Definitions
- Useful key points regarding the Greek language
- Proposed Characters for Registrations

- Issues concerned
- Proposed solutions
- Recommendations
- Appendix (table of proposed allowed characters for Greek TLD registrations)

Definitions

- Homograph
- Homophone
- Greeklish
- Aliased Name
- Name Aliasing

- Bundling Domain Names
- Tonos
- Dialytika
- Katharevousa
- Dimotiki

Useful Key Points

- ## The Greek Language Question
  Diglossia: Katharevousa vs. Dimotiki

  (e.g. Πειραιεύς - Πειραιάς)

- ## The Greek Orthography
  Polytonic vs. Monotonic (e.g. the Lord's Prayer)
  - Πάτερ ἡμῶν ὁ ἐν τοῖς οὐρανοῖς, ἁγιασθήτω τὸ ὄνομά σου, ἐλθέτω ἡ βασιλεία σου, γενηθήτω τὸ θέλημά σου, ὡς ἐν οὐρανῷ καὶ ἐπὶ τῆς γῆς
  - Πάτερ ημών ο εν τοις ουρανοίς, αγιασθήτω το όνομά σου, ελθέτω η βασιλεία σου, γενηθήτω το θέλημά σου, ως εν ουρανώ και επί της γης

ICANN

Dakar
SÉNÉGAL
N°42 23 - 28 October 2011

Proposed Characters for Registrations

Only Monotonic characters are allowed!

- Monotonic characters are used today in the spelling of Greek words.

- Polytonic characters for TLD registration offer no significant advantage for the average user.

- This recommendation stands only for the Greek TLDs; registries are encouraged to decide upon the use of the polytonic characters in their policy for the lower levels.

Issues Concerned

The sigma and final sigma

- **σ** (small, U+03C3) e.g. α**σ**πίδα (shield)
- **Σ** (capital, U+03A3) e.g. **Σ**οφία (female name)
- **ς** (small final, U+03C2) e.g. τέλο**ς** (end)

but...

- IDNA2003: **σ** <-> **Σ** <-> **ς** (e.g. θησαυρός – treasure)

  θησαυρός → ΘΗΣΑΥΡΟΣ → θησαυροσ → **not correct Greek!**
- IDNA2008: **σ** and **ς** are different accepted characters, treated separately – reverse mapping no possible!

Issues Concerned, Cont.

The IDNA approach of translating Greek words into domain names

- **.ελλάς** (.xn--hxarsa0b) small letters form
- **.ελλας** (.xn--mxahsa0b) misspelled without the tonos
- **.ΕΛΛΑΣ (.ελλασ)** (.xn--mxahsa5b) capital letters form
- ➢ in IDNA2008 uppercase letters are disallowed – <u>however</u> an application may map them to allowed characters

but...

a typical user expects to type the domain name in uppercase and lowercase and end up at the same result, as it happens in ASCII

thus:

a word (or string) without a tonos should be addressed as variant

Issues Concerned, Cont.

- **Homographs**
  - EETT (greek/capital) – EETT (latin/capital)
    εεττ (greek/small) – eett (latin/small)
  - Αθήνα – Αθηνά → ΑΘΗΝΑ

- **Homophones**
  - lotto – λοττο (or λοτο)

- **Greeklish**
  - Αθήνα → Athina / A8ina / A8hna

Issues Concerned, Cont.

- **Registrations containing the tonos accent**
  - ✓ the majority of the words in small letters have to have the tonos accent over one of the vowels to be spelled correctly
  - ✓ a domain without the tonos is consisting of the letters of the upper case form of the domain name, if the final sigma is not present
  - ✓ many domain names represent words and the user will expect them to have the tonos accent

- **Registrations containing the final sigma**
  - ✓ the use of the final sigma is essential for the correct spelling of the Greek words and male names

Issues Concerned, Cont.

What should be considered as *"variant"*?

- ✓ a variant of a label representing a word (accented by tonos in small letters' form) could be the same word in capital letters, since the capital letters are not accented,
- ✓ the variation of a given word in Katharevousa and Dimotiki,
- ✓ the monotonic and the polytonic form of the word,
- ✓ accented and not-accented forms of a word/string,
- ✓ the variation of a word/string retaining the final sigma at the exact requested position between the one when the final sigma (ς) is replaced by the middle sigma (σ),
- ✓ variations created by changing case a word/string from small to capital letters, according to the Greek grammar

Proposed Solutions – The Variants Proposal

- Accepted characters: small, monotonic, including characters with tonos, dialytika and their combination

- The domain name will be accepted at the exact requested form

- The same domain will be allocated to the registrant, stripped of the accent marks and final sigmas, if any (variant 1)

- The same domain will be allocated to the registrant, stripped of the accent marks, retaining the final sigmas, at their original positions (variant 2)

Proposed Solutions-The Variants Proposal,cont.

- Alternative positioning of the tonos accent is not allowed
- Alternative positioning of the final sigma is not allowed
- Alternative, same meaning words, are not allowed (if the domain name consists of words) – it is the "Katharevousa vs. Dimotiki" issue. The applicant will only choose to register one and the other should be excluded from registration
- Two options for handling the allocated variants:
    - they enter the root zone as DNAMEs
    - they are treated as FQDNs

Proposed Solutions – Small Characters Proposal

- Accepted characters: small, monotonic, including characters with tonos, dialytika and their combination
- The only accepted form of the domain name is the form that is originally submitted by the applicant, with the specific tonos' positions and the possible final sigmas
- The same form, stripped of accents is not allowed
- The same form, stripped of the accents and with any combination of sigma/ final sigma, is not allowed

Proposed Solutions-Small Characters Proposal,cont.

- Alternative positioning of the tonos accent is not allowed

- Alternative, same meaning words, are not allowed (if the domain name consists of words) – it is the "Katharevousa vs. Dimotiki" issue. The applicant will only choose to register one and the other should be excluded from registration

Recommendations

Taking into account:

- the advantages and disadvantages of each proposal
- the average internet user's experience, needs and expectations
- the current protocols' status and technology's limitations

The Greek Case Study Team recommends as the most appropriate solution **the Variants proposal**

Next Steps

- The Public Comment Forum on the Greek Case Study Team report ends on 14 November 2011.

- The Greek Case Study Team will analyze the comments and determine how to address them and whether to revise the Report.

- The Greek, Cyrillic and Latin Study Teams must interact to identify cross-script issues that might be present between Latin, Greek and Cyrillic characters.

Thank You

Questions

?