

Arabic Script Variant Issues for TLDs

Arabic Case Study Team



ICANN

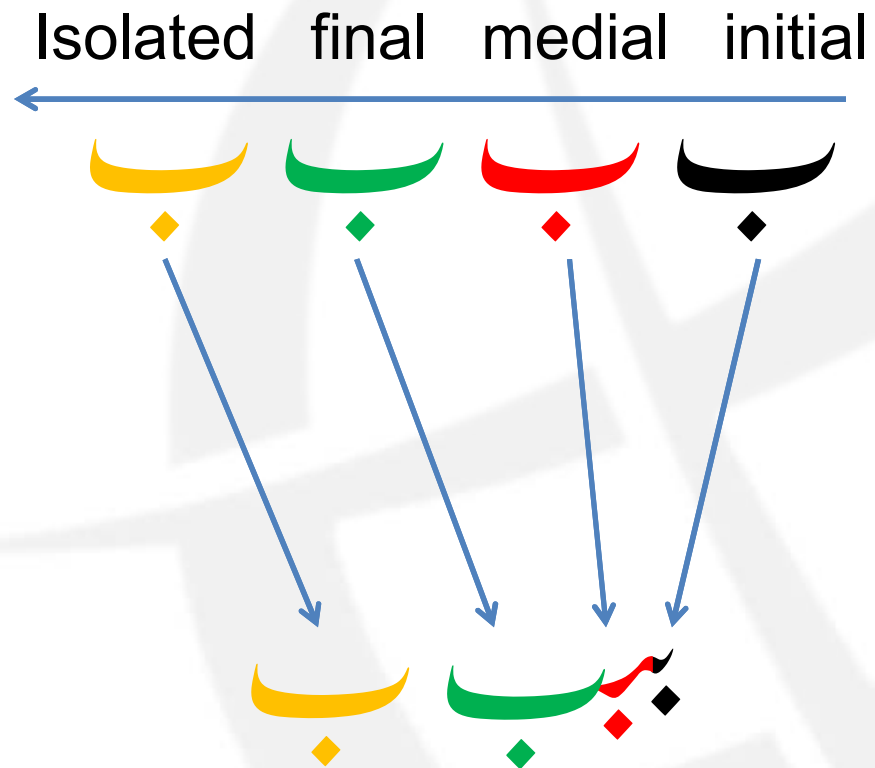
Dakar

SENEGAL

N°42 23 - 28 October 2011



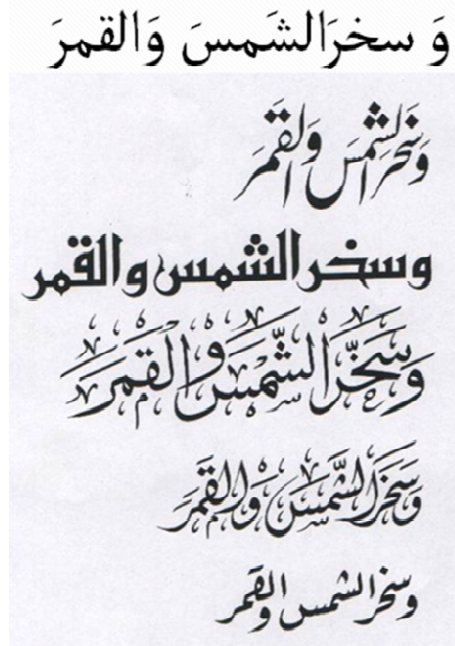
Taxonomy



Dakar
SÉNÉGAL
N°42 23 - 28 October 2011



Character Set



- Not all PVALID characters should be allowed for TLDs
 - Combining Marks
 - Archaic Characters
 - Quranic Marks
 - Digits
 - Others (misc.)



Multi-"word" Expressions

فيض آباد

فيضآباد

- Cursive complex script
- No concept of abbreviation
- Space not allowed
- No capitalization
- Possibilities discussed by the case study team – not finalized
 - ZWNJ (for some other languages)
 - Hyphen (for Arabic language)
- Neither allowed for TLDs



Variants in Arabic Script

كه ≠ كه

ن ≠ ن

كا = كا

تا = تا

آ = آ

له = له

- Identical Cases
- Similar Cases
- Interchangeable Cases
- Optional Cases



Dakar
SÉNÉGAL
N°42 23 - 28 October 2011



Label Generation Policy

Two ways for writing a letter, each with or without a combining mark

- More than a “language” “table”
- Characters + variants + additional constrains + meta information + ...
- Need an automatic variant generation and verification process
- a string of 5 letters can have 4^5 variants
- Should variant be at character or position level?



Dakar
SÉNÉGAL
N°42 23 - 28 October 2011



Variant label Status Management

- Multiple States
 - Allocated, Reserved, Blocked, Activated, Delegated, Fundamental
- Process and constraints for changing between states?
- Impact of TLD re-delegation
- Impact of change in label generation policy



WHOIS Lookup

$$\left\{ \begin{array}{c} \text{Label } 3_1 \\ \dots \\ \text{Label } 3_l \end{array} \right\} \cdot \left\{ \begin{array}{c} \text{Label } 2_1 \\ \dots \\ \text{Label } 2_m \end{array} \right\} \cdot \left\{ \begin{array}{c} \text{Label } 1_1 \\ \dots \\ \text{Label } 1_n \end{array} \right\}$$

- How will variant Domain Name Registration Data (DNRD) be looked up?
- Additional Services? Variant look up to point to Fundamental? Status information?



Fees

- Recommend no extra fees
 - Variants caused not due to applicants choice but through inherent ambiguity in
 - Script
 - Encoding
 - Increasing fees will discourage variants, affecting user access and perceived internet stability



End User Requirements/Issues

- Character support (OS, font, applications)
- Font style and size variation
- Keyboard variation
- Lack of management tools for U-label display
- Bi-directionality



Arabic Case Study Team

Local Host Organization

The Supreme Council of Information and Communication Technology (ictQatar), Qatar

List of Contributors

Sarmad Hussain	Case Study Coordinator
Abdulaziz Al-Zoman	Team Member
Behnam Esfahbod	Team Member
Fahd A. Batayneh	Team Member
Huda Sarfraz	Team Member
Iftikhar H. Shah	Team Member
Khaled Koubaa	Team Member
Manal Ismail	Team Member
Mohamed El-Bashir	Team Member
Raed Al-Fayez	Team Member
Siavash Shahshahani	Team Member
Alireza Saleh	Team Observer

ICANN staff and consultants:

Baher Esmat	Case Study Liaison
Francisco Arias	Subject Matter Expert (Registry Operations)
Kim Davies	Subject Matter Expert (Security)
Andrew Sullivan	Subject Matter Expert (Protocols)
Nicholas Ostler	Subject Matter Expert (Linguistics)



Thank You



Questions

One World

One Internet



Dakar
SÉNÉGAL
N°42 23 - 28 October 2011

