

Arabic Script Variant Issues for TLDs

Arabic Case Study Team



ICANN

Dakar

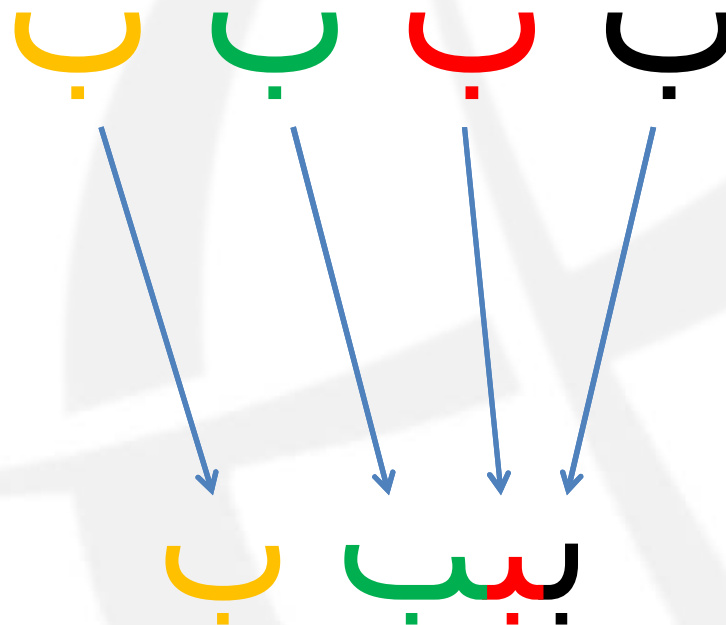
SENEGAL

N°42 23 - 28 October 2011



Taxonomy

وَ سَخِرَ الشَّمْسِ وَالْقَمَرَ
وَسَخِرَ الشَّمْسِ وَالْقَمَرَ
وَسَخِرَ الشَّمْسِ وَالْقَمَرَ
وَسَخِرَ الشَّمْسِ وَالْقَمَرَ
وَسَخِرَ الشَّمْسِ وَالْقَمَرَ
وَسَخِرَ الشَّمْسِ وَالْقَمَرَ



← Isolated final medial initial



Character Set for TLDs

	060	061	062	063	064	065	066
0	 0600	م 0610		ذ 0630	- 0640	ِ 0650	٠ 0660
1	 0601	ع 0611	ء 0621	ر 0631	ف 0641	ّ 0651	١ 0661
2	 0602	ي 0612	آ 0622	ز 0632	ق 0642	ُ 0652	٢ 0662
3	 0603	ي 0613	أ 0623	س 0633	ك 0643	ّ 0653	٣ 0663
4		ّ 0614	و 0624	ش 0634	ل 0644	ُ 0654	٤ 0664

- Not all PVALID characters should be allowed for TLDs
 - Combining Marks
 - Archaic Characters
 - Quranic Marks
 - Digits
 - Others (misc.)
- Need feedback of other language communities



Multi-Part labels

ف ي ض آ ب ا د



فيضاآباد



فيض آباد

- Cursive complex script
- Need to break the label into multiple sub-portions to meet user expectations about the writing system
 - Space not allowed
 - No capitalization
- Need mechanism(s) to enable this in labels



Dakar
S É N É G A L
N°42 23 - 28 October 2011



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

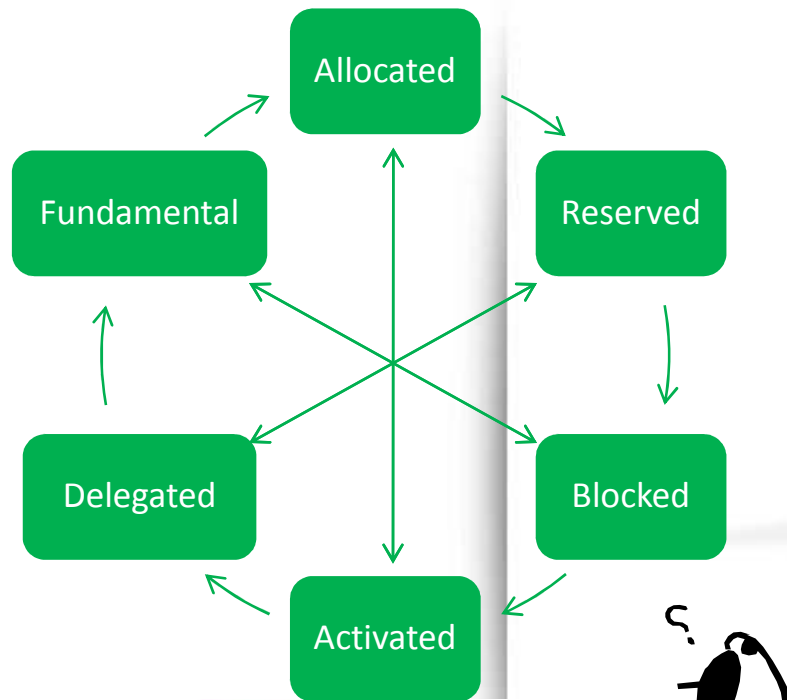
- More than a “language” “table”
- Characters + variants + additional constrains + meta information + ...
- Need an automatic variant generation and verification process
- a string of 5 letters can have 4^5 variants
- Should variant be at character or position level?



Dakar
S É N É G A L
N°42 23 - 28 October 2011



Variant label Status Management



- A label may have multiple States
 - Allocated, Reserved, Blocked, Activated, Delegated, Fundamental
- Process and constraints for assigning and changing states?
 - Impact of TLD re-delegation
 - Impact of change in label generation policy



Dakar
SÉNÉGAL
N°42 23 - 28 October 2011



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?



Dakar
SÉNÉGAL
N°42 23 - 28 October 2011



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



ICANN



Dakar

SÉNÉGAL

N°42 23 - 28 October 2011

Questions

One World

One Internet



Dakar
SÉNÉGAL
N°42 23 - 28 October 2011

