



SaudiNIC:

Securing Arabic Domain Names

Raed Alfayez, SaudiNIC APTLD74, Tashkent, Sep 2018

Agenda

- ➤ About SaudiNIC
- ➤ About Arabic Script/Language
- ➤ Securing Arabic domains
- ➤ What's Missing?

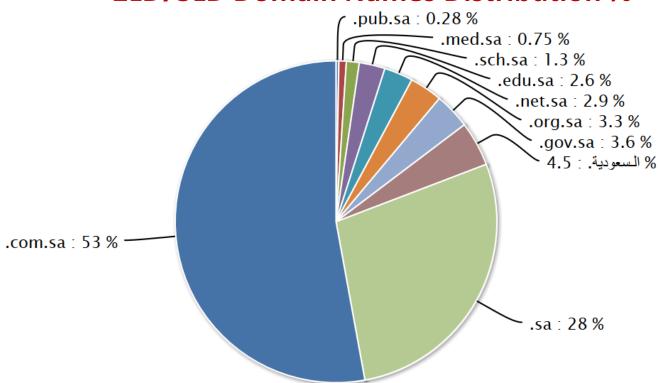
About SaudiNIC

- Administering the domain name space under:
 - (.sa) since 1995
 - (السعودية.) Experimental since 2004, official since 2010.
- Operated by a government organization:
 - CITC (Communication and Information Technology Commission)
- Leading the local and regional communities efforts towards supporting Arabic language in Domain Names since 2001 (more than 15 years of experience)
- Participated in several TF & WG related to Arabic domains regionally and globally
- Developed: Tools, Pilot Projects, Reports & RFC related to Arabic domains

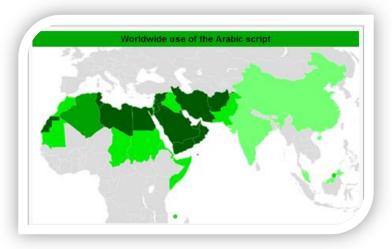
About SaudiNIC

55,850 Saudi domain names 2,540 Arabic domains under (السعودية).) 466 Variant

2LD/3LD Domain Names Distribution %



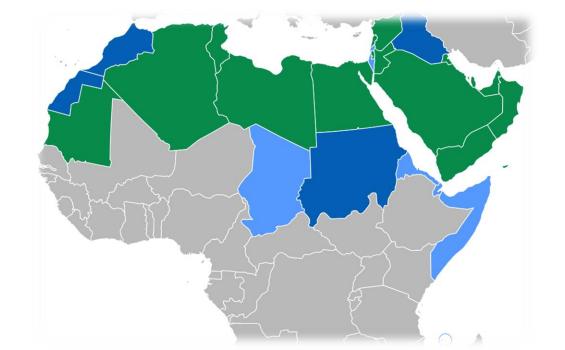
Introduction: Arabic Script



- The 2nd most widely used alphabetic writing system in the world
- Used by many languages such as:
 - Arabic, Urdu, Persian, Turkish, Kurdish, Pashto, ...etc
- It is widely used by more than 43 countries
 - more than one billion potential users could be concerned in using Arabic script domain names.

Introduction: Arabic Language

- Ranked as the 5th language by native speakers in the world.
- Considered as Official/Co-official language in 25 country, ~ 295 million native speakers.



Introduction: Arabic Language /Script

- The Arabic language is a subset of the Arabic script
 - Script: 0600..06FF, 0750..077F, 08A0..08FF Language: 0621..063A, 0641..064A, 0660..0669

L	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F		075	076	077	A80	08B	08C	08D	08E	08F
0	0600	் 0610	پ 0620	5 0630	0640	<u></u> 0650	0660	0670); 0680	0690	06A0	96B0	6	پ	0 06E0	06F0	0	<u>ي</u> 0750	9 0760	مثل 0770	پ 08A0	O8B0				(08F0
1	0601	్ 0611	چ 0621	ر 0631	ن 0641	័ 0651	0661	آ 0671	0681	ر 0691	و 06A1	06B1	06C1	ي 06D1	06E1) 06F1	1	ث 0751	پ 0761	ژ 0771	<i>ب</i> 08A1	↑ 08B1				ి 08F1
2	0602	ි 0612	T 0622	خ 0632	ق 0642	் 0652	۲ 0662	0672	0682	<u>خ</u> 0692	ب 06A2	96B2	5 06C2	06D2	○ 06E2	Y 06F2	2	ب 0752	0762	0772	<u>ٿ</u> 08A2	ĵ 08B2				9 08F2
3	<u>مب</u>	رم ا 0613	ر 0623	س 0633	<u>ئ</u> 0643	్ 0653	٣	0673	0683) 0693	ب 06A3	06B3	; 06C3	06D3	္င 06E3	₩	3	ت 0753	3 0763	Ŭ 0773	ظ 08A3					ೆ 08F3
4	 0604	ි 0614	وً 0624	ىش 0634	<u>J</u>	် 0654	ک 0664	0674	0684)	ث 06A4	06B4	_9 06C4	 06D4	○ 06E4	← 06F4	4	<u>ن</u> 0754	0764	آ 0774	ث 08A4				Ó 08E4	ే 08F4
5	0605	් 0615) 0625	ص 0635	6	ু 0655	0665	ر 0675	2 0685	پ	ب 06A5	<u>ڵ</u> 0685	9 06C5	06D5	و 06E5	△ 06F5	5	<u>ب</u> 0755	, 0765	<u>ئ</u> 0775	ق 08A5				ీ 08E5	Ő 08F5
6	0606	0616	ئ 0626	ض 0636	ن 0646	ှ 0656	0666	و 0676	0686	<u>ئ</u> 0696	ق 06A6	<u>ј</u>	<u>ۆ</u> 0606	06D6	 06E6	9 06F6	6	ٽ 0756	0766	<u>ئ</u> 0776	€ 08A6				ූ 08E6	○ 08F6
7	0607	් 0617	0627	ط 0637	<u>م</u> 0647	. 0657	V 0667	وُ 0677	0687	ن 0697	خ 06A7	ئ 06B7	ۇ 06C7	قنے ش 06D7	06E7	V 06F7	7	<u>ٿ</u>	ن ₀₇₆₇	ي 0777	څ 08A7				් 08E7	် 08F7
8	9	⊙ 0618	<u>ب</u> 0628	<u>ن</u> 0638	9 0648	് 0658	^	ځی ₀₆₇₈	5	5	څ	<u>ي</u> 0688	و 06C8		06E8	∧ 06F8	8	7.58 0758	خ 0768	5 0778	ئي _{08A8}				ි 08E8	○ 08F8
9	/	் 0619	ö 0629	0639	<u>ح</u> 0649	్ 0659	0669	ك 0679	3	ٿ 0699	06A9	ن 06B9	9 06C9	оера ,	06E9	Q	9	ے 0759	<u>ن</u> 0769	ق 0779	<u>ي</u> ₀₈ A9				ू 08E9	္ 08F9
Α	/ 060A	O61A	ت 062A	063A	<u>چ</u> 064A	Č 065A	'/- 066A	ئ 067A) 068A	بنی 069A	O6AA	ن 068A	ق 06CA	O6DA	ڼ 06EA	بث 06FA	Α	→ 075A	J 076A	077A	ک 08۸۸				O8EA	Ç 08FA
В	9 060B	6 061B	ث 062B	5 063B	О	் 065B	ر 066B	پ 067B	068B	پس _{069B}	5 06AB	ئ 06BB	ۇ 06CB	OGDB	ံ ogeb	جن 06FB	В	ر 075B	う 076B	<u>ح</u> 077B	.9 08AB				Ö 08ЕВ	ॐ 08FB
С	€ 060€	ALM 061C	062C	€	් 064C	065C	, 066C	ټ 067C	3	پژی 0690	<u>ئ</u> 06AC	ڼ _{06BC}	<u>ی</u>	OGDC	Ů 06EC	06FC	С	ش 0750	خ 0760	ر	بع 08AC				ै 08EC	ి 08FC
D	060D		O62D	<u>څ</u>	064D	े 065D	★	ٽ 067D	3 068D	<u>ص</u> 069D	<u>څ</u> 06AD	ث 06BD	ۍ 06CD	OGDD	O O O O O	اً 06FD	D	ع _{075D}	ئی 076D	س <i>ک</i> ۱) 08AD				O8ED	ै 08FD
E	060E	₽ 061E	062E	ت 063E	о 064Е	Ő 065Е	<u>ں</u> 066E	پ 067E	5	ث 069E	<u>ئ</u> 06AE	95 06BE	<u>ٽ</u> 06CE	06DE	3 06EE	(i) OGFE	Е	<u>ئ</u> 075E	076E	ش 077E	چ 08AE				OSEE	Ő 08FE
F	060F	⊆ 061F	ک 062F	ث 063F	் 064F	्र 065F	9	ت 067F	3	ظر ₀₆₉ F	06AF	O6BF	خ 06CF	ے 06DF	<u>ر</u> 06EF	ھُ 06FF	F	075F	076F	<u>خ</u> 077F	چ 08AF				ू ober	ੈ OBFF

1

Identifying allowed code points

- Identifying allowed code points at language level and/or script level
- Not allowing unsecure/unsafe code points that may effect the stability of the registry domain space (e.g. control code points, combining marks, diacritics, Tatweel ..etc).

3

No script mixing

 A label can be constructed only by Arabic code points and digits 2

Identifying variants

- Variants within a language
- Variants across the whole script
- Reachability variants: So that a registered domain name can be accessed regardless of the input devices (language table) being used by the navigator users.

4

No language mixing

(utilizing the **powerful** tools: Language tables)

- control input via the user interface
- tremendously reduce the number of unnecessary/unrealistic variants
- Registry choice to protect the TLD-space.

5

No digits mixing

- Simplify the need for 3 digits sets in Arabic script
 - European digits
 - Arabic-Indic digits
 - Eastern Arabic-Indic digits

6

WLE Rules (language/scripts)

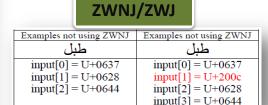
 Safeguard to handle some special cases (e.g. similarity in certain positions)

(1) Identifying code points at both language and script level

Include only:

✓ Allowed code points at language level and/or script level

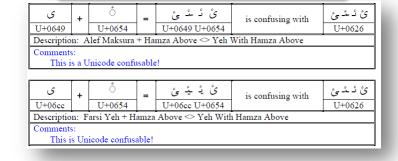




Not included:

- Combining Marks
- Diacritics (Tashkeel)
- Special codes (ZWNJ, ZWJ)
- Unused/Historical

Combining Marks

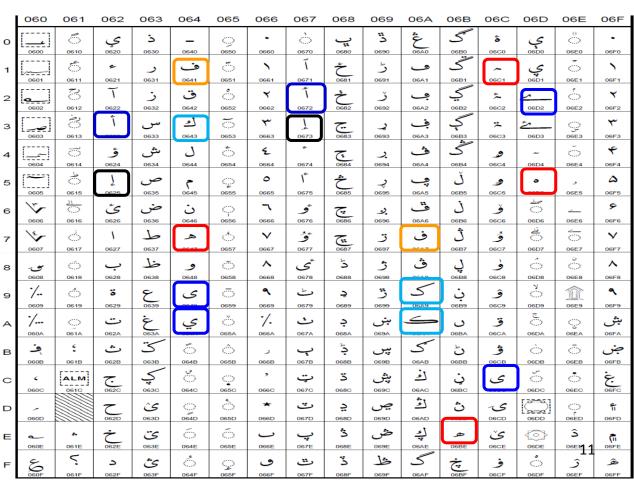


(2) Identifying variants (language level)



(2) Identifying variants (script level)

- There are a number of groups of characters that have the same shapes (Homoglyph), eg.:
 - Kaf group,
 - Heh group,
 - Yeh group,
 - Alef group
 - **–** ...



(2) Identifying variants (international reachability)

- For reachability purposes, variants should be addressed to be activated by the registry, so that:
 - A registered domain name is accessed regardless of the input devices (language table) being used by the navigator users.
 - For example:
 - A user registered the domain "مكة (all characters from the Arabic language)
 - if another user try to reach that domain name from an Internet café in Pakistan he/she will type "مكت" (all characters from the Urdu language)
 - If the "activated" variants were not allocated, delegated and hosted then the domain name will not be reachable!

Hence, reachability issue (based on input devices used by other language communities) should be carefully considered when defining variants (by language communities).



LANGUAGE	UNICODE	LABEL
	(U+0645)	1
Arabic	(U+0643)	مكة
	(U+0629)	
Persian,	(U+0645)	
Malay,	(U+06A9)	مكة
Pashto	(U+0629)	
	(U+0645)	
Urdu	(U+06A9)	مکت
	(U+06C3)	,







Securing Arabic Domains (3) No Script Mixing

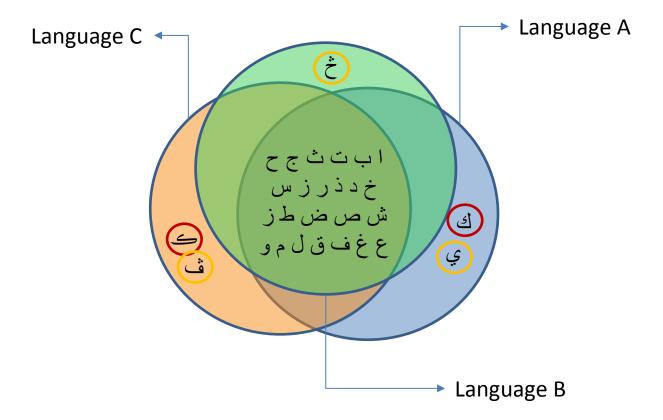
– Recommendation:

No mixing within a label or cross labels in the same domain

Format	Example
<firstname><arabic-lastname></arabic-lastname></firstname>	Salehالفلاني
<arabic-lastname><firstname></firstname></arabic-lastname>	Salehالفلاني
<tld>.<arabic-domain></arabic-domain></tld>	sa.رسیل
<domain>.<arabic-tld></arabic-tld></domain>	raseel.السعودية

(4) No language Mixing

- Not allowing code point that exists only in one language with another code points that exists only in another language in the same label
 - In the below example: 실 & = cant exist in the same label,
 - same thing for: ڦ&څ &ي



(4) No language Mixing

IDN	Total Variants	Allocatable	Blocked	Blocked due to Language Mixing			
مكة-المكرمة	3239	34	3205	3181	(99.25%)		
القرآن-الكريم	11999	111	11888	11836	(99.56%)		
هيئة-الإعلام	47999	81	47918	47764	(99.68%)		
كهف-الياسمين	28799	65	28734	28680	(99.81%)		
کهف-اکبا	21599	47	21552	21534	(99.92%)		

(5) No Digit Mixing

Digits in the Arabic scripts:

European digits	U+0030 U+0039	(0123456789)
 Arabic-Indic digits 	U+0660 U0669	(•17807779)
 Eastern Arabic-Indic digits 	U+06F0 U+06F9	(.1774897749)

#	Input string	Display
1	مؤتمر 2009	Acceptable: Pure European digits
	conference2009	
2	مؤتمر ٢٠٠٩	Acceptable: Pure Arabic-Indic digits
	conference	
3	مؤتمر ٢٠٠٩	Acceptable: Pure Eastern Arabic-Indic digits
	conference	
4	9مؤتمر ٢٠٠	Not-Acceptable: Mix between European digits & Arabic-Indic digits
	conference 7 · · 9	
5	٥٠٠٩مؤتمر	Not-Acceptable: Mix between European digits & Arabic-Indic digits
	conference2 · · ٩	
6	مؤتمر ٢٠٠٩	Not-Acceptable: Mix between Arabic-Indic digits & Eastern Arabic-
	conference	Indic digits
7	2۰۰۹مؤتمر	Not-Acceptable: Mix between European digits & Eastern Arabic-Indic
	conference2 · · ٩	digits

Securing Arabic Domains (6) WLE Rules

- HEH and TEH MARBUTA
 - They are variants only in the final form (at the end)
 - جده ۷۵ جدة –

– ALEF MAKSURA:

- Preventing connected ALEF MAKSURA
- ستر ۸۵ سر –

What's Missing?

Register and enable necessary variants:

مكة

مكة

مکت

Registry

Configure DNS & add need RRs (e.g. NS & A & CNAME) for:

xn--ogb5cf

xn--ogb9c4p

xn--hhb4rwc

DNS Hosting **Configure Email** account and email aliases:

رائد@مكة

رائد@مكة رائد@مكة

Email Hosting

Configure webserver and account and aliases:

<VirtualHost 10.10.10.10>

DocumentRoot "/makkah"

ServerName xn--ogb5cf

ServerAlias xn--oqb9c4p ServerAlias xn--hhb4rwc

VirtualHost>

Web Hosting

Automation for hosting IDN and Variants

Gift

 SaudiNIC's Best Practices in Supporting and Managing Arabic Domain Names

http://www.nic.sa/docs/SaudiNIC ADNBP.pdf

Thank you



للمزيد من المعلومات يمكنكم زيارة: For more information you can visit:



سجل.السعودية

nic.sa



هيئة-الاتصالات.السعودية citc.gov.sa