# IDNA200X Review for Arabic Script Block

Arabic Script IDN Working Group (ASIWG)
Version 1.2
16 June 2008

#### **Revision History:**

Date	Version	Changes	Changed by
28 May 2008	1.0	Created	A. Al-Zoman
9 June 2008	1.1	FE73 added, 0610-0614	Sarmad Hussain,
		removed from disallowed list	Nayyara Karamat
16 June 2008	1.2	Change in title; editing;	Sarmad Hussain
		addition of explanations and	
		captions	

This document reviews the latest output of the "IDNA200X-Tables" document that lists which codepoints are PValid, ContectO/J, UnAssigned, or Disallowed [1] for the use in Internationalized Domain Names.

The document also provides a set of recommendations by ASIWG for changes in this protocol to be considered by the IETF Internationalized Domain Names in Applications (idnabis) working group with respect to Arabic Script.

#### 1. IDNA200X Table for Arabic Script

Listed below is the list of latest output of IDNA200X-Tables (for the Arabic Script Block):

```
0600..0603 ; DISALLOWED # ARABIC NUMBER SIGN..ARABIC SIGN SAFHA
0604..0605 ; UNASSIGNED # <reserved>..<reserved>
0606..060F ; DISALLOWED # ARABIC-INDIC CUBE ROOT..ARABIC SIGN MISRA
0610..061A ; PVALID # ARABIC SIGN SALLALLAHOU ALAYHE WASSALLAM..AR
061B
          ; DISALLOWED # ARABIC SEMICOLON
061C..061D ; UNASSIGNED # <reserved>..<reserved>
061E..061F ; DISALLOWED # ARABIC TRIPLE DOT PUNCTUATION MARK..ARABIC Q
0620
           ; UNASSIGNED # <reserved>
0621..065E ; PVALID
                        # ARABIC LETTER HAMZA..ARABIC FATHA WITH TWO D
065F
          ; UNASSIGNED # <reserved>
0660..0669 ; PVALID # ARABIC-INDIC DIGIT ZERO..ARABIC-INDIC DIGIT
066A..066D ; DISALLOWED # ARABIC PERCENT SIGN..ARABIC FIVE POINTED STA
066E..0674 ; PVALID # ARABIC LETTER DOTLESS BEH..ARABIC LETTER HIG
0675..0678 ; DISALLOWED # ARABIC LETTER HIGH HAMZA ALEF..ARABIC LETTER
0679..06D3 ; PVALID # ARABIC LETTER TTEH..ARABIC LETTER YEH BARREE
           ; DISALLOWED # ARABIC FULL STOP
06D4
06D5..06DC ; PVALID # ARABIC LETTER AE..ARABIC SMALL HIGH SEEN
06DD..06DE ; DISALLOWED # ARABIC END OF AYAH..ARABIC START OF RUB EL H
06DF..06E8 ; PVALID # ARABIC SMALL HIGH ROUNDED ZERO..ARABIC SMALL
06E9
          ; DISALLOWED # ARABIC PLACE OF SAJDAH
06EA..06FF ; PVALID # ARABIC EMPTY CENTRE LOW STOP..ARABIC LETTER
FE70..FE72 ; DISALLOWED # ARABIC FATHATAN ISOLATED FORM..ARABIC DAMMAT
           ; PVALID
                        # ARABIC TAIL FRAGMENT
FE73
           ; DISALLOWED # ARABIC KASRATAN ISOLATED FORM
FE74
<FE75
          ; UNASSIGNED # <reserved
FE76..FEFC ; DISALLOWED # ARABIC FATHA ISOLATED FORM..ARABIC LIGATURE
<FEFD..FEFE ; UNASSIGNED # <reserved>..<reserved</pre>
FEFF
           ; DISALLOWED # ZERO WIDTH NO-BREAK SPACE
```

This can be reflected in the Unicode Arabic Script Block as shown in the Figure 1.

(	0600							Ara	bic							06FF
	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F
0	0600	<u>م</u> 0610		<b>3</b>	0640	O650	0660	) 0670	<b>)</b>	<b>3</b> 0690	ت 06A0	\$ 06B0	<b>6</b>	<b>ب</b>	0 06E0	06F0
1	0601	عر ٥611	<b>\$</b> 0621	ر 0631	<b>ن</b>	ឺ 0651	0661	0671	ځ	ر ا	<b>و</b>	5 06B1	~ 06C1	<b>ي</b> 06D1	う 06E1	\ 06F1
2	ل	ි	Ĩ	ز	ق	ំ	۲	ٲ	ڂ	ڒ	ڢ	2	۲.		់	۲
3	0602 [صر	ر <u>م</u> رم	0622	س 0632	<u>642</u>		0662 <b>★</b>	0672	0682	0692	06A2 <b>ب</b>	06B2	06C2 *	06D2	06E2	06F2 <b>★</b>
4	0603		وً	0633	J	0653 <b>Č</b>	0663 <b>£</b>	0673 \$	0683	), 0693	06A3 <b>ث</b>	06B3	06C3	06D3	06E3	06F3
5		0614 ط	0624	0634 ص	0644	0654 	0664	0674	0684 <b>څ</b>	0694	06A4 <b>پ</b>	06В4 Ў	06C4 <b>9</b>	06D4	06E4	06F4
6	\ <u>\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\</u>	0615	0625	<sub>0635</sub>	ر 0645	0655	0665	0675 <b>9</b>	0685	0695	ر 06A5 <b>ق</b>	06B5	06C5	06D5 ميلا	06E5	06F5
0	0606	0616	0626	0636	0646	0656	0666	0676	0686	<b>)</b> .	06A6	06B6	06C6	۰ 06D6 قع	06E6	06F6
7	0607	Ó 0617	0627	ط 0637	هر 0647	Ó 0657	<b>V</b>	<b>وُ</b> 0677	<u>ت</u> 0687	ر 0697	ف 06A7	<b>ن</b> 06B7	<b>وُ</b> 06C7	ි 06D7	06E7	<b>V</b> 06F7
8	<b>بو</b> 0608	Ó 0618	ب 0628	ظ 0638	9 0648	ŏ 0658	<b>^</b>	ځی	د 0688	ژ <sub>0698</sub>	<b>ث</b> 06A8	<u>ل</u> 0688	و 06C8	် 06D8	ن 06E8	<b>∧</b> 06F8
9	·/	Ó	ä	ع	ی	<b>্</b>	٩	ٹ	ş	ڙ	5	ڹ	ۉ	ं	Î	٩
Α	0609	0619 <u></u>	و <sub>0629</sub> ت	و 0639	و649	0659 Č	0669	<sub>0679</sub>	\$	سن (	06A9	06B9 ن	<sub>06C9</sub>	06D9 ح	06E9	 ش
В	9 19	061A	062A ث	063A	064A	065A	066A ر	067A پ	068A	069A پس	06AA	068A	06CA	06DA	06EA	06FA ض
С	060B	061B	062B	063B	064B ひ	065B	066B	067В <b>Ç</b>	068B	<sub>069B</sub>	06AB غ	06BB ڼ	06CB	06DB	06EB	06FC
	060C		062C	063C	064C	065C	066C ★	067C	068C	069C	06AC <u>څ</u>	06BC	06CC	06DC	06EC	
D	060D		062D	ئ 063D	064D	065D	066D	<b>ٿ</b> 067D	068D	<u>ص</u> 069D	06AD	ڻ 06BD	ۍ 06CD	06DD	O6ED	ÎI 06FD
Е	060E	<b>♣</b> 061E	خ	<u>ت</u>	Ó	<u>`</u>	<b>ن</b>	پ	3	ش	<u>چا</u> 06AE	<i>₽</i>	ئ	<b></b>	3	(ii
F	060E	061E	062E <b>3</b> 062F	063E	064E	065E	066F	067E	068E 3 068F	069E	JOBAE 06AF	06BE	9 06CF	06DE 0 0 06DF	06EE 3 06EF	مۇر مۇر 06FF

Figure 1(a): Arabic Script Disallowed (highlighted) and PValid/Context O/J Characters

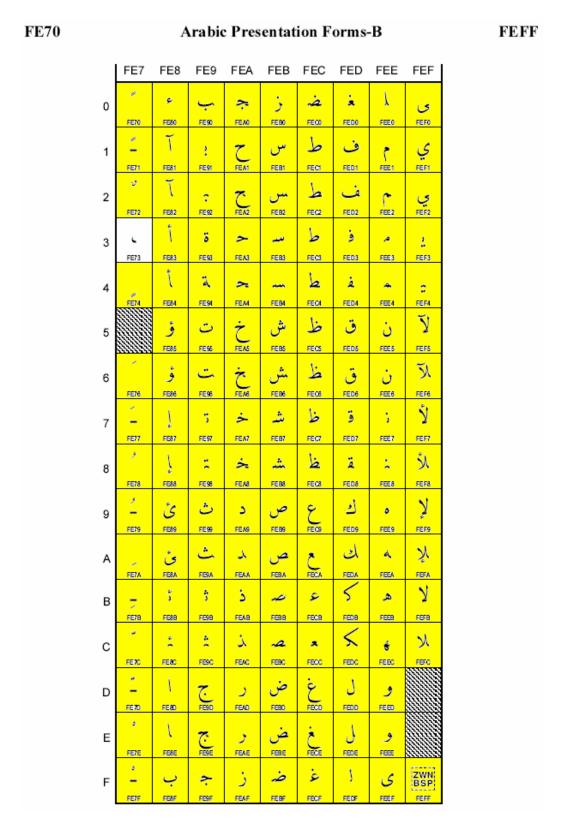


Figure 1(b): Arabic Script Disallowed (highlighted) and PValid/Context O/J Characters

### 2. ASIWG Recommended Changes

Figure 2 depicts the recommended additional code points which should be DISALLOWED by the protocol, as suggested by the ASIWG after reviewing the above list. Here is the summary of this additional recommendation:

```
0615..061A ; DISALLOWED # ARABIC SMALL HIGH TAH..ARABIC SMALL KASRA
0640 ; DISALLOWED # ARABIC TATWEEL
06D6..06DC ; DISALLOWED # ARABIC SMALL HIGH LIGTURE SAD WITH LAM WITH ALEF MAKSURA
.. ARABIC SMALL HIGH SEEN
06DF..06E8 ; DISALLOWED # ARABIC SMALL HIGH ROUNDED ZERO..ARABIC SMALL HIGH NOON
06EA..06ED ; DISALLOWED # ARABIC EMPTY CENTRE LOW STOP..ARABIC SMALL LOW MEEM
FE73 ; DISALLOWED # ARABIC TAIL FRAGMENT
```

These characters should be DISALLOWED because they are either related to stylistic variation of basic characters (e.g. Tatweel) or are Quranic characters or marks, not used in normal writing. If they are included, they may cause security issues, and their exclusion does not limit expression in any of the languages using Arabic script.

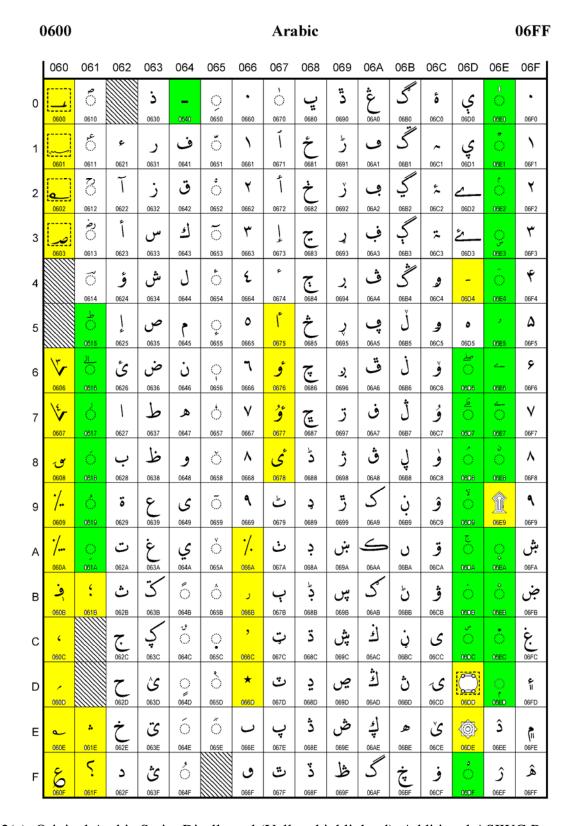


Figure 2(a): Original Arabic Script Disallowed (Yellow highlighted), Additional ASIWG Recommended Disallowed (Green highlighted), and PValid/Context O/J Characters

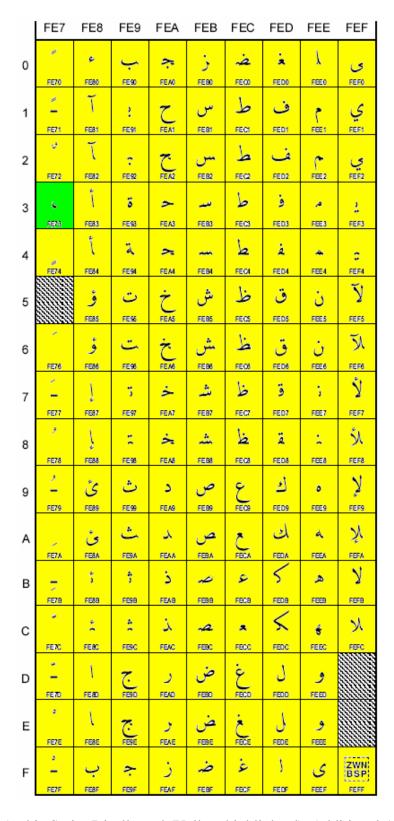


Figure 2(b): Original Arabic Script Disallowed (Yellow highlighted), Additional ASIWG Recommended Disallowed (Green highlighted), and PValid/Context O/J Characters

## 3. References

Faltstrom, P., "The Unicode Code Points and IDNA", May 5, 2008 version, available at <a href="http://tools.ietf.org/html/draft-ietf-idnabis-tables-01">http://tools.ietf.org/html/draft-ietf-idnabis-tables-01</a>.