

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 removed from disallowed list	Sarmad Hussain, Nayyara Karamat
16 June 2008	1.2	Change in title; editing; addition of explanations and captions	Sarmad Hussain

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
06D4 ; DISALLOWED # ARABIC FULL STOP
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
FE73 ; PVALID # ARABIC TAIL FRAGMENT
FE74 ; DISALLOWED # ARABIC KASRATAN ISOLATED FORM
<FE75 ; UNASSIGNED # <reserved>
FE76..FEFC ; DISALLOWED # ARABIC FATHA ISOLATED FORM..ARABIC LIGATURE
<FEFD..FEFE ; UNASSIGNED # <reserved>..<reserved>
FEFF ; DISALLOWED # ZERO WIDTH NO-BREAK SPACE

```

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

	0600	Arabic														06FF	
	060	061	062	063	064	065	066	067	068	069	06A	06B	06C	06D	06E	06F	
0	ا	آ	ب	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط
1	ع	ف	ق	ك	ل	م	ن	و	ي	ك	ل	م	ن	و	ي	ك	ل
2	هـ	و	ز	ح	ط	ي	ك	ل	م	ن	و	ي	ك	ل	م	ن	و
3	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	ف	ق	ك	ل
4	ع	ف	ق	ك	ل	م	ن	و	ي	ك	ل	م	ن	و	ي	ك	ل
5	هـ	و	ز	ح	ط	ي	ك	ل	م	ن	و	ي	ك	ل	م	ن	و
6	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	ف	ق	ك	ل
7	ع	ف	ق	ك	ل	م	ن	و	ي	ك	ل	م	ن	و	ي	ك	ل
8	هـ	و	ز	ح	ط	ي	ك	ل	م	ن	و	ي	ك	ل	م	ن	و
9	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	ف	ق	ك	ل
A	ع	ف	ق	ك	ل	م	ن	و	ي	ك	ل	م	ن	و	ي	ك	ل
B	هـ	و	ز	ح	ط	ي	ك	ل	م	ن	و	ي	ك	ل	م	ن	و
C	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	ف	ق	ك	ل
D	ع	ف	ق	ك	ل	م	ن	و	ي	ك	ل	م	ن	و	ي	ك	ل
E	هـ	و	ز	ح	ط	ي	ك	ل	م	ن	و	ي	ك	ل	م	ن	و
F	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	ف	ق	ك	ل

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

FE70

Arabic Presentation Forms-B

FEFF

	FE7	FE8	FE9	FEA	FEB	FEC	FED	FEE	FEF
0	ا	ب	ج	د	هـ	و	ز	ح	ط
1	ي	ك	ل	م	ن	س	ع	ف	ق
2	ي	ك	ل	م	ن	س	ع	ف	ق
3	ي	ك	ل	م	ن	س	ع	ف	ق
4	ي	ك	ل	م	ن	س	ع	ف	ق
5	ي	ك	ل	م	ن	س	ع	ف	ق
6	ي	ك	ل	م	ن	س	ع	ف	ق
7	ي	ك	ل	م	ن	س	ع	ف	ق
8	ي	ك	ل	م	ن	س	ع	ف	ق
9	ي	ك	ل	م	ن	س	ع	ف	ق
A	ي	ك	ل	م	ن	س	ع	ف	ق
B	ي	ك	ل	م	ن	س	ع	ف	ق
C	ي	ك	ل	م	ن	س	ع	ف	ق
D	ي	ك	ل	م	ن	س	ع	ف	ق
E	ي	ك	ل	م	ن	س	ع	ف	ق
F	ي	ك	ل	م	ن	س	ع	ف	ق

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.

	Arabic															
	0600	0601	0602	0603	0604	0605	0606	0607	0608	0609	060A	060B	060C	060D	060E	060F
0	ا	ب	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط
1	ظ	ع	ف	ق	ك	ل	م	ن	و	ى	ي	ك	ل	م	ن	و
2	ي	آ	أ	ب	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص
3	ض	ط	ظ	ع	ف	ق	ك	ل	م	ن	و	ى	ي	ك	ل	م
4	ن	و	ى	ي	ك	ل	م	ن	و	ى	ي	ك	ل	م	ن	و
5	ى	ي	ك	ل	م	ن	و	ى	ي	ك	ل	م	ن	و	ى	ي
6	ك	ل	م	ن	و	ى	ي	ك	ل	م	ن	و	ى	ي	ك	ل
7	ل	م	ن	و	ى	ي	ك	ل	م	ن	و	ى	ي	ك	ل	م
8	م	ن	و	ى	ي	ك	ل	م	ن	و	ى	ي	ك	ل	م	ن
9	ن	و	ى	ي	ك	ل	م	ن	و	ى	ي	ك	ل	م	ن	و
A	و	ى	ي	ك	ل	م	ن	و	ى	ي	ك	ل	م	ن	و	ى
B	ى	ي	ك	ل	م	ن	و	ى	ي	ك	ل	م	ن	و	ى	ي
C	ي	ك	ل	م	ن	و	ى	ي	ك	ل	م	ن	و	ى	ي	ك
D	ك	ل	م	ن	و	ى	ي	ك	ل	م	ن	و	ى	ي	ك	ل
E	ل	م	ن	و	ى	ي	ك	ل	م	ن	و	ى	ي	ك	ل	م
F	م	ن	و	ى	ي	ك	ل	م	ن	و	ى	ي	ك	ل	م	ن

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

	FE7	FE8	FE9	FEA	FEB	FEC	FED	FEE	FEF
0	س	ء	ب	ج	ز	ض	غ	ا	ي
	FE70	FE80	FE90	FEA0	FEB0	FEC0	FED0	FEE0	FEF0
1	ـ	آ	ب	ح	س	ط	ف	م	ي
	FE71	FE81	FE91	FEA1	FEB1	FEC1	FED1	FEE1	FEF1
2	ـ	آ	ب	ح	س	ط	ف	م	ي
	FE72	FE82	FE92	FEA2	FEB2	FEC2	FED2	FEE2	FEF2
3	س	أ	ة	ح	س	ط	ف	م	ي
	FE73	FE83	FE93	FEA3	FEB3	FEC3	FED3	FEE3	FEF3
4	س	أ	ة	ح	س	ط	ف	م	ي
	FE74	FE84	FE94	FEA4	FEB4	FEC4	FED4	FEE4	FEF4
5	ـ	و	ت	خ	ش	ظ	ق	ن	لا
	FE75	FE85	FE95	FEA5	FEB5	FEC5	FED5	FEE5	FEF5
6	ـ	و	ت	خ	ش	ظ	ق	ن	لا
	FE76	FE86	FE96	FEA6	FEB6	FEC6	FED6	FEE6	FEF6
7	ـ	و	ت	خ	ش	ظ	ق	ن	لا
	FE77	FE87	FE97	FEA7	FEB7	FEC7	FED7	FEE7	FEF7
8	ـ	و	ت	خ	ش	ظ	ق	ن	لا
	FE78	FE88	FE98	FEA8	FEB8	FEC8	FED8	FEE8	FEF8
9	ـ	ي	ث	د	ص	ع	ك	ه	لا
	FE79	FE89	FE99	FEA9	FEB9	FEC9	FED9	FEE9	FEF9
A	ـ	ي	ث	د	ص	ع	ك	ه	لا
	FE7A	FE8A	FE9A	FEAA	FEBA	FECA	FEDA	FEEA	FEFA
B	ـ	ي	ث	ذ	ص	ع	ك	ه	لا
	FE7B	FE8B	FE9B	FEAB	FEBB	FECB	FEDB	FEEB	FEFB
C	ـ	ي	ث	ذ	ص	ع	ك	ه	لا
	FE7C	FE8C	FE9C	FEAC	FEBC	FECB	FEDC	FECC	FEFC
D	ـ	ا	ج	ر	ض	غ	ل	و	ـ
	FE7D	FE8D	FE9D	FEAD	FEBD	FECD	FEDD	FEED	FEED
E	ـ	ا	ج	ر	ض	غ	ل	و	ـ
	FE7E	FE8E	FE9E	FEAE	FEBE	FECE	FEDE	FEFE	FEFE
F	ـ	ب	ج	ز	ظ	غ	ا	ي	ZWN BSP
	FE7F	FE8F	FE9F	FEAF	FEBF	FECF	FEDF	FEFF	FEFF

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 <http://tools.ietf.org/html/draft-ietf-idnabis-tables-01>.