

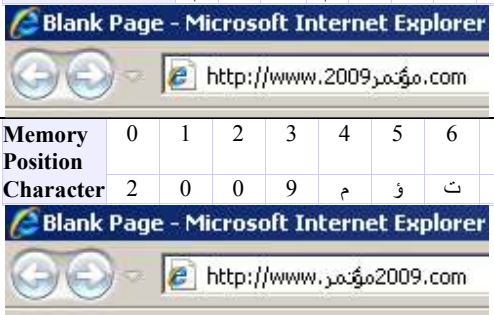
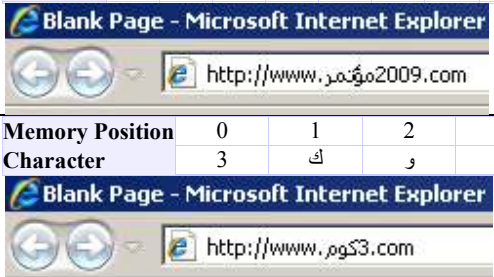

Some Concerns on the new IDNA & IDNA-Bidi Rules

Raed Al-Fayez, Abdulaziz Al-Zoman, SaudiNIC, 19 July 2009

Problem description:

After testing the current IDNA implementation on this link: <http://unicode.org/cldr/utility/idna.jsp>, we have noticed that some of the domain names that start with a Number or ends with a number (which of course always needed in domain labels) may be blocked from protocol level because of IDNA and/or IDNA-Bidi rules and the reason given was **“The first character may not be an EN (European Number) or an AN (Arabic Number)”!**

Examples:

#	Input string (Arabic language using Arabic script)	Display	English translation																				
1	مؤتمر2009	<table border="1"> <tr> <td>Memory Position</td> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td> </tr> <tr> <td>Character</td> <td>م</td><td>ؤ</td><td>ت</td><td>م</td><td>ر</td><td>2</td><td>0</td><td>0</td><td>9</td> </tr> </table> 	Memory Position	0	1	2	3	4	5	6	7	8	Character	م	ؤ	ت	م	ر	2	0	0	9	Confrances2009
Memory Position	0	1	2	3	4	5	6	7	8														
Character	م	ؤ	ت	م	ر	2	0	0	9														
2	مؤتمر2009	<table border="1"> <tr> <td>Memory Position</td> <td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td> </tr> <tr> <td>Character</td> <td>2</td><td>0</td><td>0</td><td>9</td><td>م</td><td>ؤ</td><td>ت</td><td>م</td><td>ر</td> </tr> </table> 	Memory Position	0	1	2	3	4	5	6	7	8	Character	2	0	0	9	م	ؤ	ت	م	ر	2009Confrances
Memory Position	0	1	2	3	4	5	6	7	8														
Character	2	0	0	9	م	ؤ	ت	م	ر														
3	3كوم	<table border="1"> <tr> <td>Memory Position</td> <td>0</td><td>1</td><td>2</td><td>3</td> </tr> <tr> <td>Character</td> <td>3</td><td>ك</td><td>و</td><td>م</td> </tr> </table> 	Memory Position	0	1	2	3	Character	3	ك	و	م	3Com										
Memory Position	0	1	2	3																			
Character	3	ك	و	م																			

The next figure shows the results of the test for the above “input strings”. As it can be seen, the last two input strings (#2 and #3) are denied by the protocol.

Unicode IDNA Demo

[character](#) | [properties](#) | [unicode-set](#) | [compare-sets](#) | [regex](#) | [bnf-regex](#) | [breaks](#) | [transform](#) | [bidi](#) | [idna](#) | [languageid](#) | [help](#)

Input Labels: (See also Picker)	Exclude from NFKC-CF-RDI mapping:
مؤتمـر 2009 مؤتمـر 2009 3كوم	[؁ ڄ [:Join_C:]]

[Context Rules](#)

Modified IDNA2008 Pattern

M-Label*	U-Label*	IDNA()	IDNAbis()	Error Pos.	Cause
مؤتمـر 2009	مؤتمـر 2009	<i>Denied!</i>	xn--2009-k5f0bzc5jc		
مؤتمـر 2009	مؤتمـر 2009	<i>Denied!</i>		Ⓜ مؤتمـر 2009	7. The first character may not be an EN (European Number) or an AN (Arabic Number).
3كوم	3كوم	<i>Denied!</i>		Ⓜ 3كوم	7. The first character may not be an EN (European Number) or an AN (Arabic Number).

* = Putative labels. If there are accents or invisible characters they are shown on a second line with '\u' escapes, to show the difference between cases like Ö and Ö +'

Version 3.3; ICU version: 4.2.0.0; Unicode version: 5.1.0.0

Done

To run the same test, please click on the following link:

[http://unicode.org/cldr/utility/idna.jsp?a=%D9%85%D8%A4%D8%AA%D9%85%D8%B12009%0D%0A2009%D9%85%D8%A4%D8%AA%D9%85%D8%B1%0D%0A3%D9%83%D9%88%D9%85&f=\[%C3%9F+%CF%82+\[%3AJoin_C%3A\]\]](http://unicode.org/cldr/utility/idna.jsp?a=%D9%85%D8%A4%D8%AA%D9%85%D8%B12009%0D%0A2009%D9%85%D8%A4%D8%AA%D9%85%D8%B1%0D%0A3%D9%83%D9%88%D9%85&f=[%C3%9F+%CF%82+[%3AJoin_C%3A]])

Recommendations:

The new IDNAbis() should be revised to allow the use of digits at the beginning of a label.