IDN Tables for ‫رازاب.‬

Table of Contents

IDN Table for the Arabic Script

TLD: بازار.

Script Tag: Arabic

Version: 1.0

Effective Date: 2012-03-16

Registry: dotبازار (dotBazaar) Registry

For the development of this IDN Table, CORE Internet Council of Registrars was fortunate to have as team members Siavash Shahshahani and Alireza Saleh, two experienced IDN experts who have together been involved in both the technical and policy aspects of IDNs, especially in Arabic script. They have participated in a number of ICANN working groups, IETF for IDNA2003 and IDNA2008, and various other activities. They are also responsible for the development of Persian IDN at .IR Registry (IPM/IRNIC) as represented in IANA IDN Repository http://www.iana.org/domains/idn-tables/tables/ir_fa_1.0.html. CORE Internet Council of Registrars also plans to seek the advice of other experts from major language groups that may be served by the proposed TLD.

For every registered Arabic script IDN, an index (or canonical) string will be computed and stored in the SRS database. The index string is produced by applying the rules from the section below to the IDN U-label. If no rule applies or if all applying rules map their respective character to itself, then the index string will be the same as the registered string. Once an Arabic script IDN is registered, other domain names that map to the same index string will be unavailable for registration. Additionally, there are some rules that specifically prohibit a certain combination of Arabic characters within a domain label. Such labels will also be unavailable for registration.

The table presented here is in compliance with the ICANN Guidelines for the Implementation of Internationalized Domain Names Version 3.0 and is intended for publication in the IANA Repository of TLD IDN Practices, for the information of prospective holders of domains in بازار. and for the users of resources within those domains.

1. Table of Allowed Characters and their Joining Property

The table below lists the characters allowed in the Unicode representation of IDNs associated with the Arabic script. The second column shows the character property used in the rules below. Columns are delimited by semicolons. The "#" symbol denotes start of a comment that continues to the end of line.

```
# Valid code point ; character property
U+002D; U # HYPHEN-MINUS
U+0030; U # DIGIT ZERO
U+0031; U # DIGIT ONE
U+0032; U # DIGIT TWO
U+0033; U # DIGIT THREE
U+0034; U # DIGIT FOUR
U+0035; U # DIGIT FIVE
U+0036; U # DIGIT SIX
U+0037; U # DIGIT SEVEN
U+0038; U # DIGIT EIGHT
U+0039; U # DIGIT NINE
U+0621; U # ARABIC LETTER HAMZA
U+0622; R # ARABIC LETTER ALEF WITH MADDA ABOVE
U+0623; R # ARABIC LETTER ALEF WITH HAMZA ABOVE
U+0624; R # ARABIC LETTER WAW WITH HAMZA ABOVE
U+0625; R # ARABIC LETTER ALEF WITH HAMZA BELOW
U+0626; D # ARABIC LETTER YEH WITH HAMZA ABOVE
U+0627; R # ARABIC LETTER ALEF
U+0628; D # ARABIC LETTER BEH
U+0629; R # ARABIC LETTER TEH MARBUTA
U+062A; D # ARABIC LETTER TEH
U+062B; D # ARABIC LETTER THEH
U+062C; D # ARABIC LETTER JEEM
U+062D; D # ARABIC LETTER HAH
U+062E; D # ARABIC LETTER KHAH
U+062F; R # ARABIC LETTER DAL
U+0630; R # ARABIC LETTER THAL
U+0631; R # ARABIC LETTER REH
U+0632; R # ARABIC LETTER ZAIN
U+0633; D # ARABIC LETTER SEEN
U+0634; D # ARABIC LETTER SHEEN
U+0635; D # ARABIC LETTER SAD
U+0636; D # ARABIC LETTER DAD
U+0637; D # ARABIC LETTER TAH
U+0638; D # ARABIC LETTER ZAH
U+0639; D # ARABIC LETTER AIN
U+063A; D # ARABIC LETTER GHAIN
U+0641; D # ARABIC LETTER FEH
U+0642; D # ARABIC LETTER QAF
U+0643; D # ARABIC LETTER KAF
U+0644; D # ARABIC LETTER LAM
U+0645; D # ARABIC LETTER MEEM
U+0646; D # ARABIC LETTER NOON
U+0647; D # ARABIC LETTER HEH
U+0648; R # ARABIC LETTER WAW
U+0649; D # ARABIC LETTER ALEF MAKSURA
U+064A; D # ARABIC LETTER YEH
U+0660; U # ARABIC-INDIC DIGIT ZERO
U+0661; U # ARABIC-INDIC DIGIT ONE
U+0662; U # ARABIC-INDIC DIGIT TWO
U+0663; U # ARABIC-INDIC DIGIT THREE
U+0664; U # ARABIC-INDIC DIGIT FOUR
U+0665; U # ARABIC-INDIC DIGIT FIVE
U+0666; U # ARABIC-INDIC DIGIT SIX
U+0667; U # ARABIC-INDIC DIGIT SEVEN
U+0668; U # ARABIC-INDIC DIGIT EIGHT
U+0669; U # ARABIC-INDIC DIGIT NINE
U+0679; D # ARABIC LETTER TTEH
U+067E; D # ARABIC LETTER PEH
U+067F; D # ARABIC LETTER TEHEH
U+0686; D # ARABIC LETTER TCHEH
U+0688; R # ARABIC LETTER DDAL
U+0690; R # ARABIC LETTER DAL WITH FOUR DOTS ABOVE
U+0691; R # ARABIC LETTER RREH
U+0695; R # ARABIC LETTER REH WITH SMALL V BELOW
U+0698; R # ARABIC LETTER JEH
U+0699; R # ARABIC LETTER REH WITH FOUR DOTS ABOVE
```

```
U+069C; D # ARABIC LETTER SEEN WITH THREE DOTS BELOW AND THREE DOTS ABOVE
U+069E; D # ARABIC LETTER SAD WITH THREE DOTS ABOVE
U+06A0; D # ARABIC LETTER AIN WITH THREE DOTS ABOVE
U+06A2; D # ARABIC LETTER FEH WITH DOT MOVED BELOW
U+06A4; D # ARABIC LETTER VEH
U+06A5; D # ARABIC LETTER FEH WITH THREE DOTS BELOW
U+06A7; D # ARABIC LETTER QAF WITH DOT ABOVE
U+06A8; D # ARABIC LETTER QAF WITH THREE DOTS ABOVE
U+06A9; D # ARABIC LETTER KEHEH
U+06AE; D # ARABIC LETTER KAF WITH THREE DOTS BELOW
U+06AF; D # ARABIC LETTER GAF
U+06B5; D # ARABIC LETTER LAM WITH SMALL V
U+06BA; D # ARABIC LETTER NOON GHUNNA
U+06BD; D # ARABIC LETTER NOON WITH THREE DOTS ABOVE
U+06BE; D # ARABIC LETTER HEH DOACHASHMEE
U+06C1; D # ARABIC LETTER HEH GOAL
U+06C6; R # ARABIC LETTER OE
U+06C7; R # ARABIC LETTER U
U+06C8; R # ARABIC LETTER YU
U+06CA; R # ARABIC LETTER WAW WITH TWO DOTS ABOVE
U+06CC; D # ARABIC LETTER FARSI YEH
U+06CE; D # ARABIC LETTER YEH WITH SMALL V
U+06CF; R # ARABIC LETTER WAW WITH DOT ABOVE
U+06D2; R # ARABIC LETTER YEH BARREE
U+06D5; R # ARABIC LETTER AE
U+06F0; U # EXTENDED ARABIC-INDIC DIGIT ZERO
U+06F1; U # EXTENDED ARABIC-INDIC DIGIT ONE
U+06F2; U # EXTENDED ARABIC-INDIC DIGIT TWO
U+06F3; U # EXTENDED ARABIC-INDIC DIGIT THREE
U+06F4; U # EXTENDED ARABIC-INDIC DIGIT FOUR
U+06F5; U # EXTENDED ARABIC-INDIC DIGIT FIVE
U+06F6; U # EXTENDED ARABIC-INDIC DIGIT SIX
U+06F7; U # EXTENDED ARABIC-INDIC DIGIT SEVEN
U+06F8; U # EXTENDED ARABIC-INDIC DIGIT EIGHT
U+06F9; U # EXTENDED ARABIC-INDIC DIGIT NINE
U+0762; D # ARABIC LETTER KEHEH WITH DOT ABOVE
U+200C;   # ZERO WIDTH NON-JOINER
```

2. Pattern Syntax

The following rules will have a pattern describing a character from the above table to be matched. Since the Arabic characters cannot be considered on a one-by-one basis, a look-behind and look-ahead pattern is (optionally) added to each rule. A rule applies to a domain label if the given pattern matches a character within the label and the characters before and after match the patterns given in look-behind and look-ahead, respectively. The terms 'before' and 'after' are considered to be in relation to the on-the-wire order of the characters.

The following character properties will be used in the patterns.

- **{L}** denotes characters that are left-joining
- **{R}** denotes characters that are right-joining
- **{D}** denotes characters that are dual-joining (i.e. both left- and right-joining)
- **{U}** denotes characters that are unable to join
- **{T}** denotes combining characters

- **{1}** denotes the ASCII digits Zero - Nine (i.e., 0030..0039)
- **{2}** denotes the Arabic digits Zero - Nine (i.e., 0660..0669)
- **{3}** denotes the extended Arabic digits Zero - Nine (i.e., 06F0..06F9)
- character classes can be combined, e.g. **{LD}** denotes characters that are either left-joining or dual-joining

Apart from the character properties necessary for the Arabic characters, the used pattern syntax is defined by the following syntax elements which are based on the regular expression syntax.

- **.** denotes any character
- **^** denotes the beginning of the label
- **$** denotes the end of the label
- **|** denotes an alternative
- **(...)** groups a subpattern
- an asterisk **(*)** denotes any number of occurrences, including zero
- **ZWNJ** denotes the zero width non-joiner character (U+200C)
- **U+XXXX** denotes the Unicode character with the hexadecimal code point XXXX

3. Rules Describing Invalid Labels

If any of the following rules is matched, the respective domain label is rejected and may not be registered. Note that some special cases of the rules below are already included in IDNA2008. Excluding those special cases would only have made the rules more complicated and less intuitive so they were kept.

```
Comment:       prevent confusion that may arise in conjunction with certain fonts
Look-behind:
Pattern:       U+0637 | U+0638 | U+06BE
Look-ahead:    ZWNJ {T}* {RD}
Action:        reject
```

```
Comment:       a label may not start with a digit
Look-behind:   ^
Pattern:       {123}
Look-ahead:
Action:        reject
```

```
Comment:       consecutive hyphens are not allowed in a label
Look-behind:   U+002D
Pattern:       U+002D
Look-ahead:
Action:        reject
```

```
Comment:       no mixing of the three digit sets is allowed (part 1)
```

```
Look-behind:  {23} .*
Pattern:      {1}
Look-ahead:
Action:       reject
```

```
Comment:      no mixing of the three digit sets is allowed (part 2)
Look-behind:  {13} .*
Pattern:      {2}
Look-ahead:
Action:       reject
```

```
Comment:      no mixing of the three digit sets is allowed (part 3)
Look-behind:  {12} .*
Pattern:      {3}
Look-ahead:
Action:       reject
```

4. Rules for Variant and Index Generation

If any of the following rules matches, the character from the pattern will have an index character. This index character is used to determine the index string for the whole label (simply by replacing each character by its index character). Furthermore, labels with at least one matching rule will have label variants. The variants are determined by replacing the character by any one of the allowed variants.

Note that for every variant, the previous rules to determine invalid labels also have to be checked to make sure a variant is a valid domain label.

4.1 YEH Group

```
Comment:      YEH Group (part 1)
Look-behind:
Pattern:      U+064A | U+06CC
Look-ahead:   {T}* {RD}
Index:        U+064A
Variants:     U+064A, U+06CC
```

```
Comment:      YEH Group (part 2)
Look-behind:
Pattern:      U+0649 | U+06CC
Look-ahead:   {T}* {U} | $
Index:        U+0649
Variants:     U+0649, U+06CC
```

4.2 HEH Group

```
Comment:      HEH Group
```

```
Look-behind:
Pattern:      U+0647 | U+06BE | U+06C1 | U+06D5
Look-ahead:
Index:        U+0647
Variants:     U+06A7, U+06BE, U+06C1, U+06D5
```

4.3 NOON Group

```
Comment:      NOON Group
Look-behind:
Pattern:      U+0646 | U+06BA
Look-ahead:   {T}* {RD}
Index:        U+0646
Variants:     U+0646, U+06BA
```

4.4 KEH Group

```
Comment:      KEH Group
Look-behind:
Pattern:      U+0643 | U+0649
Look-ahead:   {T}* {RD}
Index:        U+0643
Variants:     U+0643, U+06A9
```

4.5 FEH Group

```
Comment:      FEH Group
Look-behind:
Pattern:      U+0641 | U+06A7
Look-ahead:   {T}* {RD}
Index:        U+0641
Variants:     U+0641, U+06A7
```

4.6 PEH Group

```
Comment:      PEH Group
Look-behind:
Pattern:      U+067E | U+06BD
Look-ahead:   {T}* {RD}
Index:        U+067E
Variants:     U+067E, U+06BD
```

4.7 "6A0" Group

```
Comment:      6A0 Group (part 1)
Look-behind:  {LD} {T}*
Pattern:      U+06A0 | U+06A4 | U+06A8
```

```
Look-ahead:    {T}* {RD}
Index:         U+06A0
Variants:      U+06A0, U+06A4, U+06A8
```

```
Comment:       6A0 Group (part 2)
Look-behind:   ^ | {U} {T}*
Pattern:       U+06A4 | U+06A8
Look-ahead:    {T}* {RD}
Index:         U+06A4
Variants:      U+06A4, U+06A8
```

4.8 Digits

```
Comment:       Digit Zero
Look-behind:
Pattern:       U+0030 | U+0660 | U+06F0
Look-ahead:
Index:         U+0030
Variants:      U+0030, U+0660, U+06F0
```

```
Comment:       Digit One
Look-behind:
Pattern:       U+0031 | U+0661 | U+06F1
Look-ahead:
Index:         U+0031
Variants:      U+0031, U+0661, U+06F1
```

```
Comment:       Digit Two
Look-behind:
Pattern:       U+0032 | U+0662 | U+06F2
Look-ahead:
Index:         U+0032
Variants:      U+0032, U+0662, U+06F2
```

```
Comment:       Digit Three
Look-behind:
Pattern:       U+0033 | U+0663 | U+06F3
Look-ahead:
Index:         U+0033
Variants:      U+0033, U+0663, U+06F3
```

```
Comment:       Digit Four
Look-behind:
Pattern:       U+0034 | U+0664 | U+06F4
Look-ahead:
Index:         U+0034
Variants:      U+0034, U+0664, U+06F4
```

```
Comment:       Digit Five
Look-behind:
Pattern:       U+0035 | U+0665 | U+06F5
Look-ahead:
Index:         U+0035
Variants:      U+0035, U+0665, U+06F5
```

```
Comment:      Digit Six
Look-behind:
Pattern:      U+0036 | U+0666 | U+06F6
Look-ahead:
Index:        U+0036
Variants:     U+0036, U+0666, U+06F6
```

```
Comment:      Digit Seven
Look-behind:
Pattern:      U+0037 | U+0667 | U+06F7
Look-ahead:
Index:        U+0037
Variants:     U+0037, U+0667, U+06F7
```

```
Comment:      Digit Eight
Look-behind:
Pattern:      U+0038 | U+0668 | U+06F8
Look-ahead:
Index:        U+0038
Variants:     U+0038, U+0668, U+06F8
```

```
Comment:      Digit Nine
Look-behind:
Pattern:      U+0039 | U+0669 | U+06F9
Look-ahead:
Index:        U+0039
Variants:     U+0039, U+0669, U+06F9
```