

# Explanatory Annex

## 1. Generalities

- The IDN Table contains 104 characters comprised of 72 letter-characters, three sets of digits, hyphen-minus and the demarcation character ZWNJ.
- Writing is from right to left, but numbers with more than one digit are written left to right.
  - An Arabic letter character may have up to four different shapes of occurrence ('cases') which we call initial, medial, terminal and isolated, determined as follows: Initial if not joined to the character immediately to the right but joined to the character immediately to the left.
  - Medial if joined to both characters on the immediate right and left.
  - Terminal if joined to the character immediately to the right but not joined to the character immediately to the left.
  - Isolated if not joined to either of the characters immediately to the right or left.
- All characters are classified as to the type of joining they admit. A right-joining character (R) is one that can join to the character immediately to the right but NOT to the character immediately to the left. A dual joining character (D) can join to both characters on the left and right. Certain characters do not join on either side (U), e.g., digits and hyphen. There are no left-joining characters in the Arabic script, i.e., characters that join ONLY on the left.
- By definition, a right-joining character is NEVER joined on the left. But in many languages that use the Arabic script, a dual-joining character MAY or MAY NOT join on the left, often creating two strings with different connotations. The absence of join in these strings calls for a special device since no empty spacing within a label is allowed in DNS. The device, known as the zero-width-non-joiner (ZWNJ), is actually a half-space to the left of the character, and it carries the coding U+200C in UNICODE. The use of ZWNJ, however, must follow certain restrictions that will be delineated in the next section.
- There are two sets of digits in Arabic script, Western Arabic Digits (U+0660-9) and Eastern Arabic (U+06F0-9). The appearance of digits in the two sets differs only for digits corresponding to 4, 5 and 6. The UNICODE bidirectional property of the two digit sets is also different. Further, there is widespread use of ASCII digits along with Arabic letters. A special case is that of Jawi, an official alphabet for Malay, based on Arabic script. ASCII digits are used in Jawi for numerals; however the digit 'two' (<٢> in both Western and Eastern Arabic) is used as a special marker in Jawi. The occurrence of this character indicates the repetition of the syllable immediately preceding it.

## 2. Rules of label formation under the TLD

- The length of a label is limited so as its conversion into ASCII punycode does not exceed 63 ASCII characters.
- The initial character of a label must be a letter character.
- Consecutive hyphens are not allowed in a label.
- The use of ZWNJ follows IDNA2008 guidelines. This is adopted from Unicode Consortium Guideline on Layout and Format Control Characters. See: [http://www.unicode.org/reports/tr31/#Layout\\_and\\_Format\\_Control\\_Characters/](http://www.unicode.org/reports/tr31/#Layout_and_Format_Control_Characters/)
- Thus, ZWNJ may be used to the left of a D but only if followed by a D or R.
- In addition, the following restriction will also be observed regarding the use of ZWNJ: This mark may not be used following any of the three characters <U+0637>, <U+0638> or <U+06BE> followed by U. This is to prevent confusion that may arise in conjunction with the use of certain fonts.
- No mixing of the three digit sets is allowed in a label.
- All other IDNA2008 rules and restrictions apply.

## 3. Variants

- Within the framework of dotBazaar Character Table (Question 15a), variants are defined with reference to individual constituent characters in a label.
- Corresponding digits in the three digit sets will be considered variants.
- Characters considered as variants are listed in Variant Table (Question 15a). These variants are either visually identical or are used interchangeably by native users.
- Two labels will be called variants if they have the same length and their corresponding characters are either the same or are variants.
- For each legally constructible label in dotBazaar Character Table, the variant set (v-set) consists of all the variants of the label (constructed according to Variant Table). The label with the lowest UNICODE value in the v-set is designated as the v-index and of the label.
- Upon the application for a label as second-level domain, the v-index for the label is automatically generated unless it already exists by virtue of an earlier registration of a label that has the same v-index. The following situations can occur:
  - In the absence of prior existence of the v-index, the applied for label is registered (assuming other

### Q15A Explanatory Annex

requirements for registration are met), and the v-index is kept for future reference. The labels in the v-set will be banned for future registration except by this applicant, but this applicant can register other labels in the v-set ONLY through the same registrar.

- If the v-index of an applied-for label exists, indicating prior registration of a variant, the label cannot be registered except by the original registrant, as indicated above.
- It will not be possible to make a partial transfer of a subset of labels in a single v-set to a new holder. In the event one or more members of a v-set are transferred to a new (single) holder, any remaining previously registered labels in the v-set will be automatically de-registered.
- For the purpose of dispute resolution, all registered variants in a single v-set will be treated as a single domain.

## 4. Whois

If a member of a v-set is queried from the WHOIS server, and that label is not registered while some other member of the same v-set is, then the registered label(s) of the v-set will be given to the user as the result. The result will also indicate that while the queried string has not been registered, it is not available for registration except by the registrant of the previously registered members of the v-set.