

Dear IAB

With disappointment, we read your [Statement on Identifiers and Unicode 7.0.0](#) that was published on your website.

We DO understand the problem that was highlighted in your statement (pertaining the confusability of the code point (U+08A1) as it has no normalization in Unicode). However, we were astonished and shocked by your conclusions and recommendations at the end of your statement as it suggested to exclude some characters and character sequences from use in any new identifiers (all from the Arabic script) even if these characters (e.g., U+0623, U+0624, and U+0626) have some normalization rules and behave similarly to the behavior of the LATIN SMALL LETTER A WITH DIAERESIS (U+00E4), as shown in your statement.

On the other hand, your statement did not recommend excluding the Latin characters (i.e., U+00E4) or the sequence LATIN SMALL LETTER A (U+0061) followed by COMBINING DIAERESIS (U+0308) from use in any new identifiers even if they behave similarly to some of the characters that you recommended to be excluded.

It would be more realistic, rational, and workable deduction to us if your statement were concerned ONLY on the Non-spacing combining mark characters (e.g., U+0654) that are used to produce some characters without having normalization forms for them (e.g., U+08A1) regardless of the script.

Your statement as it is now (if not changed or modified) is a damaging statement to the IDN development and progress that have been made for the last decade. There are a number of TLD operators (e.g., *السعودية ، إمارات ، قطر ، مصر ، عمان ، موقع ، شبكة ، بازار*) that are currently using some of these characters in domain names. I hope you are aware of the fact that these characters that were suggested to be excluded from use in any new identifiers by your statement (e.g., U+0623, U+0624, and U+0626) are VERY ESSENTIAL to the Arabic script based languages and these languages will become unusable if these characters were excluded.

To illustrate that these characters are safe to be used in identifiers, they (e.g., U+0623, U+0624, and U+0626) are used daily as identifiers (e.g., user IDs and/or passwords for bank accounts, and domain names for websites) without any problems. BTW, with respect to domain names, the code point (U+0677), which you recommended to be excluded, is already DISALLOWED by the IDNA 2008 protocol.

As we do not have control over what the IDNA2008 protocol (RFC 5892) chooses to place a code point as PVALID or DISALLOWED, nevertheless, we as the Arabic script based language community has recommended not to use Non-Spacing Combining Marks (see: IDN Variant TLD Program Reports: [Arabic Case Study Team Report](#), *Section 5. TLD Label Valid Code Points for Arabic Script, page 4, dated 7 Oct 2011*). Here is some text from the report (concerning the recommendations related to combining marks, Unicode 5.1):

1. 0610-061A: an issue as they are PVALID but **should not be allowed** for TLDs as these are **combining marks**
2. ... [text deleted] ...
3. ... [text deleted] ...

4. 064B-0659: an issue as they are PVALID but **should not be allowed** for TLDs as these are **combining marks**
5. 065A-065F: an issue as they are PVALID but **may not be allowed** for TLDs as these are **combining marks**
6. ... [text deleted] ...
7. ... [text deleted] ...
8. 0670: an issue as they are PVALID but **should not be allowed** for TLDs as it is a **combining mark**
9. ... [text deleted] ...
10. 0674: an issue as it is PVALID but resembles a **combining mark**
11. ... [text deleted] ...
12. ... [text deleted] ...
13. 06D6-06DC: an issue as they are PVALID but **should not be allowed** for TLDs as they are Quranic marks which are not used in writing contemporary Arabic script based languages and are **combining marks**
14. 06DF-06E8: an issue as they are PVALID but **should not be allowed** for TLDs as they are Quranic marks which are not used in writing Arabic script based languages and are **combining marks**
15. 06EA-06ED: an issue as they are PVALID but **should not be allowed** for TLDs as they are Quranic marks which are not used in writing Arabic script based languages and are **combining marks**
16. ...

Additionally, the [Arabic Case Study Team Report](#) concluded by stating the following (page 9):

"A general rule may be extracted that combining marks should not be allowed for TLDs."

Regrettably, we do feel disappointed from your statement that came as a surprise to us without prior community consultations. However, we are sure (after the above reasoning) that you (IAB) will reconsider your statement by correcting its conclusions (expanding the warning for all affected scripts and limiting the excluded code points to Non-Spacing Combining Marks) or withdraw it until you get the facts straight out.

Sincerely yours
Abdulaziz H. Al-Zoman