

Feedback on PRI 343 (Unicode Emoji 5.0)

Authors: Roozbeh Pournader (Android Text team, Google Inc.)

Date: January 24, 2017

Purpose

This document provides feedback on revision 10 of the draft document for Unicode Emoji 5.0, found at <http://www.unicode.org/reports/tr51/tr51-10.html> and dated 2016-12-07.

This feedback was reviewed in the emoji ad hoc meeting on January 24, 2017, and there was general consensus that they should be applied to

1 Default emoji presentation for emoji tag sequences

Definition ED-14a in Section 1.4.5 contains a review note about default emoji presentation for emoji tag sequences. We think emoji tag sequences should *not* have default emoji presentation if the their tag base is not default emoji. Here are some of the reasons:

1. There are already two existing mechanisms for overriding the emoji presentation of default-text characters: the emoji variation selector (VS16), and emoji modifiers. Both of these need a table lookup of at most two characters (the base character + either the VS or the modifier) to decide if the available emoji font supports the sequence or not, in order to decide if we should use a text font instead. Changing that to a sequence of arbitrary length requires both arbitrary lookup into later in the string and passing arbitrary strings down to the font system in order to determine if a sequence is supported. This can make text processes unnecessarily complicated, if not almost impossible with certain architectures.
2. The fallback as specified, with any tag character following any default-text emoji character making the base display in emoji style practically creates a million different ways for making the same emoji (assuming the font does not support putting question marks over emoji). Just add any unsupported tag sequence instead of VS16. This would create headaches for filtering and searching operations which need to find certain emoji in text.

Instead, we suggest that whenever a tag sequence is used with default-text tag base character (and there is no modifier), VS16 should be present to clarify this, and help with font selection and other text processes.

In the light of this, at the emoji ad hoc meeting mentioned above, it was suggested and agreed that in the light of this, since the base character proposed in the current draft, U+1F3F3

WAVING WHITE FLAG, is *not* default-emoji, it's better to change the base glyph for the flag tag sequences to U+1F3F4 WAVING BLACK FLAG.

2 Mixing ZWJ sequences and tag sequences

A review note under Definition ED-17 calls for a mix of ZWJ sequences and tag sequences. We are very concerned about the complexity that would create and we wish to ask for keeping the emoji syntax simpler rather than complicate it.

3 three-digit region subtags

Since none of the 3-digit region subtags seem to have a recognizable flag, we think they should be removed from the syntax in Annex C. Because the sequences with such 3-digit region subtags would still be syntactically correct (though currently invalid), they can be added later if needed, without any compatibility issues.

Acknowledgments

Mark Davis and Markus Scherer reviewed the document and provided very useful technical and editorial feedback that are incorporated in the final version.

References

1. Mark Davis and Peter Edberg. "Proposed Update Unicode Technical Report #51: Unicode Emoji", Version 5.0 Revision 10, dated 2016-12-07.
<http://www.unicode.org/reports/tr51/tr51-10.html>