

Re:	Which characters should have emoji-style by default? (V2)
To:	UTC
From:	Mark Davis
Live:	http://goo.gl/zUQrff

I had an action to “Coordinate an ad-hoc committee to look at emoji style default issues and variation selectors.” For various reasons I wasn’t able to really push this along, I did refine the strawman discussed in committee, and have the following proposal for further discussion.

Requirement

In unicode.org/Public/UNIDATA/StandardizedVariants.html we have variation sequences for different characters, although not for all of those that could be expected to take an emoji style.

However, what we do not have is some sense of which characters typically have, by default, an emoji representation. We are starting to see many emoji presentations pop up, but without consistency across platforms. (Note that the EmojiSource.txt is not a reliable guide to whether a character has a default emoji presentation, because we’ve added many other similar characters, and similar characters existed already.)

That means that a piece of text may show up in a different style than intended when shared across platforms. While this is all a perfectly legitimate for Unicode—*we don’t guarantee presentation style*—it would be useful to have a more shared sense among developers of when to use emoji presentation by default, so that there are fewer “jarring” presentations.

Strawman Proposal

Here is a strawman proposal for discussion, based on the spreadsheet at <http://goo.gl/XHbvxZ>.

1. We add a new informative binary property `Emoji_Default` (ed). Where in column #2 (“Proposed 7.0 Default”) has the value “emoji”, the characters in column 1 are given the value **Emoji_Default=Yes**.
2. Where in column #3 (“Proposed 7.0 SVs”) there is the value “SV”, the characters in column 1 are given standardized variation sequences of “text style” and “emoji style” (where they don’t already have that in v6.3).

Some rows have more than one codepoint in the first column. The only relevant ones are the ‘flag’ sequences. The characters `U+1F1E6..U+1F1FF` would be given `Emoji_Default=Yes` and standardized variation sequences of “text style” and “emoji style”.

Other Information

The spreadsheet contains other columns of information for comparison:

1. Age
2. Character (for when you have a font that displays the character).
3. 6.3 SVs: SV if the character has a standard variation sequence in 6.3
4. 6.3 J Sources: the values from the EmojiSources.txt
5. Names: the character names
6. Block: the block for the character

I’ve found it useful to copy the data and paste as text into a program that displays emoji (eg TextEdit on the Mac) to see what the default behavior is on a given system.

Strategy

Here is the strategy use for a first cut of these values. I suggest that an ad hoc group review the assignments and refine where necessary.

Emoji_Default: include all those that had by default an emoji presentation in either Apple or Google fonts, plus similar characters (many in 7.0). There was then some tweaking.

SV: retain all the ones in v6.3 (of course), and add sequences for those characters that were not emoji by default, and not marked “text-only”.