

Reactivate UTS 52 mechanism in reduced form

Date: 2016-Aug-3

To: UTC

From: Peter Edberg, Mark Davis, and emoji ad-hoc committee

This is a proposal to reactivate a portion of the tag mechanism originally proposed in [draft UTS 52](#), but in modified form and with reduced scope, with initial support only for subregion flags and for private use tags. The modifications are intended to provide improved fallback on systems that do not support the tag sequences. The reduction in scope is motivated by the fact that a different mechanism (with better fallback) is proposed for gender variants, while the direction and hair color tags in the original proposal were deemed of lower priority than the other proposed tags.

One option is to reactivate UTS 52 to describe the new reduced mechanism. Alternatively, we could add it to UTR 51, which has the advantage of “non-stop” shopping.

Here are the changes proposed to the syntax portion of [UTS 52 draft 3](#)

Overall syntax - changes

The customization syntax uses the 95 invisible TAG characters:

U+E0020..U+E007F (TAG SPACE..CANCEL TAG).

These correspond to ASCII characters, and may be referenced by an abbreviated name of the form Tag-`<single-ascii-character>`, such as **Tag-F for U+E0046 TAG LATIN CAPITAL LETTER F**. In examples, where clear, they can also be represented simply by the corresponding underlined ASCII letters. The tag-term can be represented by **◆**. The regex characters `?`, `*`, `+` have their normal meaning.

In addition, there are the following special terms:

Notation	Characters	Discussion
tag-base	emoji character, emoji variation sequence, emoji modifier sequence	Expanded to include emoji_modifier_sequence
tag-term	(see discussion below)	terminating Tag ◆
tag-keyChar	Tag-A..Tag-Z	Tag characters corresponding to uppercase letters: <u>A</u> .. <u>Z</u>
tag-valChar	U+E0020..U+E0040, U+E005B..U+E007E	Tag characters that are neither tag-term nor tag-keyChar.

For best fallback on older systems, the tag-term should be a visible character, with default emoji presentation, that provides an indication or warning that some information is missing. It should also be a less commonly used character, to minimize conflict with other usages. The recommended character is:

? U+2753 BLACK QUESTION MARK ORNAMENT (distinctive on all platforms - colorful or badged).
Emojitracker rank 347, frequency = .002 x frequency of (most common)

Other characters considered for tag-term were:

! U+2757 HEAVY EXCLAMATION MARK SYMBOL (colorful on all platforms). Emojitracker rank 153,
frequency = .011 x frequency of

✗ U+274C CROSS MARK (red on all platforms, has the sense in English of “extension”, but may also have
the undesirable sense of negating the previous character). Emojitracker rank = 145, frequency = .012 x
frequency of

On systems that support specific tag sequences, the tag-term character would be absorbed as part of the
sequence, and not separately displayed.

ED-16a. emoji tag sequence (add to ED-17 emoji sequence)

A sequence consisting of an emoji character followed by one or more non-terminating TAG characters,
followed by tag-term.

```
emoji-tag-sequence      :=      tag-base tag-key-value-pair+ tag-term
tag-key-value-pair      :=      tag-key tag-value
tag-key                  :=      tag-keyChar+
tag-value                :=      tag-valueChar+
```

(no other definitions needed)

Flags - changes

Syntax

Tag-Base	U+1F3F3 WAVING WHITE FLAG
Tag-Key	Tag-F
Tag-Value	(Tag-0..Tag-9, Tag-a..Tag-z)+

...

Example: <U+1F3F3>Fgbsct✦ requests a flag for the subdivision “gbsct”, which represents Scotland. Note that
there is no hyphen, and it is all lowercase, unlike the format for ISO subdivisions (“GB-SCT”).

Gender attribute - replaced by ZWJ mechanism proposed elsewhere

Hair attribute - removed pending further study

Direction attribute - removed pending further study

Private Use - no changes

Response to relevant feedback on earlier draft UTS 52

Per PRI 321, feedback was collected on the earlier draft of UTS 52, in <http://www.unicode.org/review/pri321/feedback.html>

Most of that feedback pertains to the tags for gender, direction, and hair color in that draft of UTS 52, and is not relevant to the current proposal a more limited version of UTR 52. The feedback that is relevant, and the response by the emoji subcommittee from discussions before the May UTC meeting, is given below:

1. **2016 Mar 06 feedback from Andrew Main. Response:**
 - a. UTC is deliberately repurposing the formerly-deprecated TAGs, so no new set is needed.
 - b. The UTC has already taken care of segmentation properties for TAGs.
 - c. The proposal no longer uses the CANCEL TAG for termination.
2. **2016 Apr 04 feedback from Agustin Fonts:**

Response:

 - a. Consider an informative list of what subregion flags are commonly supported, like the catalog of ZWJ sequences.
3. **2016 Apr 19 feedback from Derek Giromini:**

Response:

 - a. See feedback on item 2 above.