Title: Emoji vs Terminals vs East_Asian_Width

Source: Ken Lunde (Adobe)

Status: Individual contribution

Reference: Action Item 145-A121

Action: For consideration by the UTC

Date: 2016-01-21

Proposal

Change the East_Asian_Width property value from A (*East Asian Ambiguous*) or N (*East Asian Neutral*) to W (*East Asian Wide*) for the 825 characters with the UTR #51 (*Unicode Emoji*) property value *Emoji_Presentation*, and add a recommendation to Section 5 of UAX #11 (*East Asian Width*), Recommendations, to the effect that "emoji style" variation sequences be treated as though their East_Asian_Width property values were W (*East Asian Wide*).

The Problem

This document is in response to Action Item 145-A121 ("Action Item for Mark Davis, Peter Edberg, Emoji SC: Investigate Phil Armstrong's error report from Oct 2, 2015. See document L2/15-277.").

For convenience, Mr. Armstrong's error report is shown below in its entirety:

Date/Time: Fri Oct 2 12:53:10 CDT 2015

Name: Phil Armstrong

Report Type: Error Report

Opt Subject: Emoji should be East Asian Full Width?

I note (from playing around with wcswidth()) that emoji characters are single/half width.

Since emoji originate from Japanese text display formats, shouldn't they be marked as East-Asian Full Width so that they will take up two columns on monospace terminal displays? As it stands, the glyphs overlap adjacent characters on monospace terminals, because they are not marked as being full-width, despite clearly being designed to take up more horizontal space than a monospace character.

cheers, Phil

In other words, should emoji whose East_Asian_Width property value is N (*East Asian Neutral*) or A (*East Asian Ambiguous*) be changed to W (*East Asian Wide*) so that they do not display in an overlapping fashion in terminals, which have a strong tendency to enforce a monospaced text layout?

First Scope Issue: General Versus East Asian Usage

Section 2 of UAX #11, Scope, is as follows:

> *The East_Asian_Width is an informative property and provides a useful concept for implementations that*
>
> - *Have to interwork with East Asian legacy character encodings*
> - *Support both East Asian and Western typography and line layout*
> - *Need to associate fonts with unmarked text runs containing East Asian characters*
>
> *This annex gives general guidelines how to use this property. It does not provide rules or specifications of how this property might be used in font design or line layout, because, while a useful property for this purpose, it is only one of several character properties that would need to be considered. While the specific assignments of property values for given characters may change over time, it is generally not intended to reflect evolving practice for existing characters. In particular some alphabetic and symbol characters are treated as wide in certain East Asian legacy character set implementations, and as narrow in all other cases. Instead, the guidelines on use of this property should be considered recommendations based on a particular legacy practice that may be overridden by implementations as necessary.*

The scope issue is whether the display of emoji in terminals is considered general or East Asian usage. While it is true that the initial set of emoji originated from Japanese carriers that continued the East Asian convention of displaying non-ASCII characters as wide (full-width) characters, the scope of emoji has arguably outgrown such scope.

Some terminals, such as Apple's [OS X Terminal app](#), have an option to force East Asian Ambiguous characters to be wide, but no such option exists for East Asian Neutral characters, at least in this particular app.

Terminal developers, or perhaps the developers of the libraries that they use, such as ones that includes *wcswidth()* that was cited by Mr. Armstrong, seem to use East_Asian_Width property values for general use, for the lack of a better or more appropriate property value. The use of East_Asian_Width property values for general use could also be construed as borderline abuse.

In defense of Mr. Armstrong's suggestion, most characters with the UTR #51 property value *Emoji_Presentation* display better as wide characters as opposed to narrow ones.

Second Scope Issue: What Is An Emoji?

A total of 838 characters are assigned the UTR #51 property value *Emoji_Presentation* ("A character that, by default, should appear with an emoji presentation, rather than a text presentation"). Of these, 13 have the East_Asian_Width property value W (*East Asian Wide*) and thus have no display issues, 800 have the East_Asian_Width property value N (*East Asian Neutral*), and 25 have the East_Asian_Width property value A (*East Asian Ambiguous*). I prepared a [file](#) that shows all of these 838 characters, organized by East_Asian_Width property value, which can be used to demonstrate the reported behavior in terminals.

The scope issue is whether to limit the scope of any East_Asian_Width property value changes to characters whose UTR #51 property value is *Emoji_Presentation*, which directly feeds into this proposal. Also within this scope are "emoji style" variation sequences, though their treatment is best handled as a recommendation.

Alternative Solutions

Below are two alternative solutions to Mr. Armstrong's error report:

1. Do nothing, on the grounds that displaying emoji in terminals is considered to be outside the scope of UAX #11, and thus someone else's problem.

2. Completely rethink the way in which East_Asian_Width property values are assigned, particularly for emoji and emoji-like characters that tend to display better as wide characters in monospaced environments. However, this solution is likely to involve an overhaul of UAX #11 that will affect its scope and intent.

That is all.