Re: UCD Emoji Properties?
From: Mark Davis
Date: 2015-010-31
Draft: https://goo.gl/ix0S9V

On the unicode.org list, there was a question as to why the emoji data was not in Unicode properties. That is a valid question, because we have given implementers a structure for relating UCD data files to properties, and how to use them (eg in regex).

Ken answered, then I added:

> As Ken said, there's been some preliminary discussion, but we wanted to get initial information out in connection with UTR #51 first, and take more time to consider what UCD properties would look like, and which are necessary.
>
> The basic information that people want to access for implementations are:
> 1. Is a character emoji or not?
> 2. Which emoji have default text presentation? (others having emoji presentation)
> 3. Which emoji are modifiers, and which are modifier bases? (others being neither)
> 4. Which sequences of emoji are recommended (zwj and/or combining marks) for those who support them?
> 5. flags and modifier sequences are specified algorithmically, and don't need to be listed.
>
> The levels, the distinction between primary and secondary, and the carrier sources were useful in development of the emoji data and tr51 but aren't really necessary for implementations.

Proposal

Add 4 binary character properties, described instead UTR #51. These would not be *UCD* properties, but would share the same namespace and structure. We would adapt the emoji-data.txt v2.0 structure for these.

Emoji (EM)
- Yes for the characters in emoji-data.txt

Emoji_Presentation (EP)
- Yes for the emoji presentation characters in emoji-data.txt

Emoji_Modifier (EM)
- Yes for the 5 Emoji Modifier characters in emoji-data.txt

Emoji_Modifier_Base (EMB)
- Yes for the current primary+secondary in emoji-data.txt