Technical Reports

L2/14-093

Working Draft Unicode Technical Report #51

UNICODE EMOJI

Version	1.0 (draft 3)
Editors	Mark Davis (markdavis@google.com), Peter Edberg (pedberg@apple.com)
Date	2013-03-21
This Version	http://www.unicode.org/draft/reports/tr51/tr51.html
Previous Version	*n/a*
Latest Version	n/a
Latest Proposed Update	n/a
Revision	1

Summary

This document provides background information about Emoji characters in Unicode, plus data to help implementations determine which characters can or should be displayed with an emoji-style presentation, rather than a text-style.

Status

*This is a **working draft** document which may be updated, replaced, or superseded by other documents at any time. Publication does not imply endorsement by the Unicode Consortium. This is not a stable document; it is inappropriate to cite this document as other than a work in progress.*

Please submit corrigenda and other comments with the online reporting form [Feedback]. Related information that is useful in understanding this document is found in the References. For the latest version of the Unicode Standard see [Unicode]. For a list of current Unicode Technical Reports see [Reports]. For more information about versions of the Unicode Standard, see [Versions].

Contents

1 Introduction
2 Presentation Style
3 Labels
4 Background
5 Data Files

Acknowledgments
References
Modifications

---

1 Introduction

ROUGH DRAFT!

Emoji are "picture characters" originally associated with cellular telephone usage in Japan, but also used in other East Asian countries and in other contexts. Use of emoji is growing quickly outside East Asia; emoji on smartphones and in chat and email applications are becoming popular. Here are some examples of recent news about emoji (as of this writing):

- The Colbert Report: *Emoji Ethnicity*
- The Wall Street Journal: *Emoji Origins*
- The Verge: *Emoji invades Twitter on the web*
- Wired: *Game of Thrones Fans, Here's Your Season Three Recap — In Emoji*
- Huffington Post: *Google Chrome Prank Translates Every Single Word Into Emoji*
- Marketplace (public radio): *You can now search Yelp for emojis*
- Huffinton Post: *You Can Now Use Emojis To Search On Yelp, And It's Not As Pointless As It Sounds*
- BuzzFeed: *This Real-Time Emoji Tracker Will Blow Your Mind*
- The Voice Online: *Company creates world's first black emojis for messaging*
- CNET Japan: *Carriers unifying on Unicode Emoji* (machine-translated English version)

[Review Note: These are useful to give context during development. But we might remove them or move them elsewhere if we think they'll go stale.]

Emoji are often pictographs—images of things such as faces, weather, vehicles and buildings, food and drink, animals and plants—or icons that represent emotions, feelings, or activities. Many emoji characters are presented in color (sometimes as a multicolor image), and some are presented in animated form, usually as a repeating sequence of two to four images—for example, a pulsing red heart. For more information, see the Unicode Emoji and Dingbats FAQ.

In early 2007, the Unicode technical committee decided to form a subcommittee for symbols (including emoji). The main goal was to allow interoperability with the many emoji characters with widespread use on Japanese mobile phones. Pictographs had long been present in Unicode since 1993, but the the first emoji-like characters in Unicode were added for interoperability with the ARIB set in 2009 with version 5.2. The largest group of emoji were then added in 2010 with version 6.0. The correspondence to the Japanese carrier symbols is in a data file EmojiSources.txt. A few more emoji were added in 2012 with version 6.1, and a large number are being added in 2013 with version 7.0.

The longer-term goal for implementations should be to support embedded graphics.

That would allow arbitrary emoji symbols, and not be dependent on additional Unicode encoding. An example of where this is done is Captain America Skype Emoji. However, until SMS, chat, mobile phones, email programs all support embedded graphics (with quick palettes of emoji-style images), implementations will often need to use plain-text Unicode emoji instead.

The goal of this document is to provide guidance in the use of emoji,

[Review Note: some of all of the data presented here might be better off in the Unicode CLDR project instead.]

2 Presentation Style

Certain emoji have defined variation sequences. These variation sequences can be used to request either an emoji presentation, or a text (black & white) presentation. For more information on these selectors, see the file StandardizedVariants.html. Implementations should support both styles of presentation for these characters, if possible. Most of those characters were those emoji that were unified with preexisting characters. Because people are now using emoji presentation for broader set of characters, it is anticipated that more such variation sequences are needed.

It is not clear for implementations what the default presentation for pictographs should be: emoji or text? 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 characters—presentation style is never guaranteed—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. That is, implementations need to be able to know what the generally expected default presentation is, to promote interoperability across platforms and applications.

[Review Note: where a character could reasonable be used with either presentation, we should propose variation sequences for it.]

3 Labels

Emoji are not typically typed on a keyboard. Instead, they are generally picked from a palette, or recognized via a dictionary. The palettes need to be organized in a meaningful way for users, which cannot be provided by either the Unicode code point order, or the Unicode Collation ordering. They typically provide a small number of broad categories (5-10), such as People (anything associated with people), Nature, and so on. These categories typically have 100-200 emoji.

For dictionary recognition, and for search, it is useful to provide annotations for emoji. These are much finer grained keywords. When you type in "hourglass" on your mobile phone, you could see and pick from either of ⌛ or ⏳.

For both palette categories and annotations, there is no requirement for uniqueness: an emoji should show up wherever users would expect them. A gas pump ⛽ might show up under "object" and "travel"; a heart 💔 under "heart" and "emotion".

There is one further kind of label, called a "read-out", for text-to-speech. For

accessibility when reading text, it is useful to have a semi-unique name for an emoji. A Unicode character name can often serve as a basis for this, but its requirements for uniqueness often ends up with names that are overly long, such as *black right-pointing double triangle with vertical bar* for ⏯.

Note that the labels need to be in each user's language to useful. They cannot simply be a translation of an English label, since different words, or even different categorizations, may be what is expected in different languages. The terms given in the data files here have been collected from different sources. They are only initial suggestions, not expected to be complete, and only in English.

4 Background

[Review Note: People seem to be interested in knowing more about the history and usage of emoji, so we're planning to add some informative material here.]

5 Data Files

This is a working draft document, and the data is supplied for now in HTML files, so that people can see sample glyphs for the characters. The available files are:

- full-emoji-list.html – the main file: a list with images showing depictions from different sources, and the default status and labels.
    - *See below for details about this file.*
- emoji-data.txt – a plaintext file with the information from the html file, plus the ordering. For now, the U+ is present, to make importing into a spreadsheet easier.
- missing-emoji-list.html – a list with images showing where sources don't have emoji images.
    - The images are *not* what would appear in that source; instead, they show cases that are marked *missing* for that source in the full-emoji-list file. So, for example, the image of ☎ in the Android column means that that character (U+260E *black telephone*) is marked as *missing* for Android in full-emoji-list. Characters in a 'common' row are missing in all of the sources: the image of 🇨🇦 there means that *all* the sources are missing the Canadian flag.
- emoji-list.html – an abbreviated list showing characters, not images. For checking browser/platform support.
- emoji-style.html – the proposed default presentation style for each character. Separate rows show the presentation with and without variation selectors, where applicable. Flags are shown with images. *Also in column 6 of the main file.*
- emoji-labels.html – characters grouped by palette category. (These are building blocks for palettes, which generally group them together.) *Also in column 7 of the main file.*
- emoji-annotations.html – characters grouped by annotation (more granular than the labels). *Also in column 7 of the main file.*
    - Clicking on an image goes to the respective row in the full-emoji-list.
- other-labels.html – other general symbols and punctuation. That can be used to scan for other characters that might qualify for emoji presentation.

- emoji-ordering.html – draft ordering of emoji characters that groups like characters together. *Unlike the labels or annotations, each character only occurs once.*

These are all live documents and may be updated or changed at any time during development.

For the full-emoji-list file, the columns are:

1. **Count.** A line count, for reference.
2. **Code.** The code point(s) for the emoji characters. Some rows have more than one codepoint where a sequence is required, such as for flags and keycaps.
    - Clicking on the code point a link to that row in the address bar.
3. **Browser.** The plaintext character, showing whatever image would be native for the browser.
4. **B&W.** The visual appearance of the codes, using the Symbola font, plus PNGs for the flags.
5. **Apple, Android, Twitter, Windows**. Low resolution images from the respective sources for comparison.
    - Note that for the cells marked *missing*, there are sometimes B&W images that would appear on the source that are not shown here. For example, U+2639 ☹ is shown as *missing* for Apple, but there are B&W images for it available on Apple platforms. Such cases should be fixed in a future version of these charts.
6. **Name.** The character name in lowercase (or an informative gloss, for the case of flags and keycaps).
7. **Version.** The version of Unicode in which the emoji was added (or will be, for Unicode 7.0). A * indicates that the character is in the Japanese sources.
8. **Default.** The draft proposed default presentation style. A * indicates that there are variation selectors (text and emoji) for the character.
9. **Annotations.** A rough-draft list of informative annotations. Clicking on a link goes to the respective row in the emoji-annotations.

Each image has the respective character as an alt value, so copying the image into plain text should (OS permitting) give the plain text character for that image. The Symbola font can be installed for a readable text presentation where the emoji presentation or black&white fonts are not available on your

Other files show the same data from different perspectives, for comparison. These files just use the characters, not images. The Symbola font can be useful for viewing them. Hovering over the character shows the code point and name. Your browser zoom is also useful.

Some historical documents used in the development of Unicode emoji may be useful, since they show the original Japanese images and the first proposed reference glyphs.

- emojidata.html - searchable doc, but the images for DoCoMo and SoftBank are currently blocked
- emojidata.pdf - large download.

[Review Note: We would like feedback such as:

1. Characters that should be added to this list, or removed from it
    - removal would be if the character is really never suited for use in an emoji presentation
2. Suggested changes to the draft proposed default presentation
    - changing the default from emoji to text or vice versa
3. Suggested changes to the labels: additions, removals, or replacements
    - This is not the main focus of the document, and may not be retained in a final version
4. Suggested additional characters that should be given variation sequences
    - eg, to allow people to request one presentation or another

]

Acknowledgments

TBD.

References

TBD

...	...
[Unicode]	The Unicode Standard *For the latest version, see:* http://www.unicode.org/versions/latest/
[UTR36]	UTR #36: *Unicode Security Considerations* http://www.unicode.org/reports/tr36/
[UTS39]	UTS #39: *Unicode Security Mechanisms* http://www.unicode.org/reports/tr39/
[Versions]	Versions of the Unicode Standard http://www.unicode.org/versions/ *For details on the precise contents of each version of the Unicode Standard, and how to cite them.*

Modifications

The following summarizes modifications from the previous revisions of this document.

Revision 1

- Working draft based on committee discussions
- Added draft data files (as HTML for viewing)
- Updated text, changed files to use images for viewing across platforms.
