L2/14-280

Working Draft **Unicode Technical Report #51**

UNICODE EMOJI

Version	1.0 (working draft)
Editors	Mark Davis (markdavis@google.com), Peter Edberg
Date	2014-10-15
This Version	http://www.unicode.org/reports/tr51/tr51-1d.html
Previous Version	*n/a*
Latest Version	n/a
Latest Proposed Update	n/a
Revision	1

Summary

The main goal of document is to help improve the interoperability of emoji characters across implementations by providing guidelines and data.

The guidelines include design recommendations with guidance for improving interoperability across platforms and implementations and longer-term approaches to emoji. Background information about emoji is also supplied.

The data includes information about Unicode emoji characters, including: which characters normally can be considered to be emoji; which of those should be displayed by default with a text-style versus an emoji-style; how to sort emoji characters more naturally; useful categories for character-pickers for mobile and virtual keyboards; and useful annotations for searching emoji.

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*This is a **working draft** document which may be updated, replaced, or superseded by other documents at any time.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

WORKING DRAFT!

Emoji are pictographs (pictoral symbols) that are typically presented in a colorful cartoon form and used inline in text. They represent things such as faces, weather, vehicles and buildings, food and drink, animals and plants, or icons that represent emotions, feelings, or activities. Emoji on smartphones and in chat and email applications have become popular worldwide.

The word *emoji* comes from the Japanese:

絵 (e ≅ picture) 文 (mo ≅ writing) 字 (ji ≅ character).

Within text, emoji may be represented internally by images or as encoded characters. When necessary for disambiguation, the latter may be called *emoji characters*. Some characters are normally displayed as emoji; some are normally displayed as ordinary text, and some can be displayed both ways.

Emoji became available in 1999 on Japanese mobile phones. There was an early proposal in 2000 to encode DoCoMo emoji in Unicode. At that time, it was unclear whether these characters would come into widespread use—and there wasn't support from the Japanese carriers to add them to Unicode—so no action was taken.

The emoji turned out to be quite popular in Japan, but each mobile phone carrier developed different (but partially overlapping) sets, and each mobile phone vendor used their own—incompatible—text encoding extensions. The vendors developed

cross-mapping tables to allow limited interchange of emoji characters with phones from other vendors, including email. Characters from other platforms that could not be displayed were represented with 〓 (U+3013 GETA MARK), but it was all too easy for the characters to get corrupted or dropped.

When non-Japanese email and mobile phone vendors started to support email exchange with the Japanese carriers, they ran into those problems. Moreover, there was no way to represent these characters in Unicode, which was the basis for text in all modern programs. In 2006, Google started work on converting Japanese emoji to Unicode private-use codes, leading to the development of internal mapping tables for supporting the carrier emoji via Unicode characters in 2007.

There are, however, many problems with a private-use approach, and thus a proposal was made to the Unicode Consortium for expanding the scope of symbols to encompass emoji. This proposal was approved in May of 2007, leading to the formation of a symbols subcommittee, and in August of 2007 the technical committee agreed to support the encoding of emoji in Unicode based on a set of principles developed by the subcommittee. The following are a few of the documents tracking the progression of Unicode emoji characters.

Date	Title	Authors
2000-04-26	NTT DoCoMo Pictographs	Graham Asher (Symbian)
2006-11-01	Symbols (scope extension)	Mark Davis (Google)
2007-08-03	Working Draft Proposal for Encoding Emoji Symbols	Kat Momoi, Mark Davis, Markus Scherer (Google)
2007-08-09	Symbols draft resolution	Mark Davis (Google)
2007-09-18	Japanese TV Symbols (ARIB)	Michel Suignard (Microsoft)
2009-01-30	Emoji Symbols Proposed for New Encoding	Markus Scherer, Mark Davis, Kat Momoi, Darick Tong (Google); Yasuo Kida, Peter Edberg (Apple)
2009-03-05	Proposal for Encoding Emoji Symbols	
2010-04-27	Emoji Symbols: Background Data (also emojidata.html; missing DoCoMo/SoftBank images)	
2011-02-15	Wingdings and Webdings Symbols	Michel Suignard

The first Unicode characters explicitly intended as emoji were actually added for interoperability with the ARIB set in 2009, with Unicode 5.2. For the Japanese mobile phone carriers, a set of 722 characters was defined as the union of their emoji characters; of these, 114 were mapped to characters already in Unicode 5.2, and the remaining 608 characters were added in 2010, with Unicode 6.0. Several other emoji characters were added to Unicode at the same time. A few more pictographs were added in 2012 with Unicode 6.1, and a large number were added in 2014, with Unicode 7.0.

Here is a summary of when some of the major sources of pictographs used as emoji were encoded in Unicode:

Source	Abbr	Dev. Starts	Released	Unicode Version	Sample Character			
					B&W	Color	Code	Name

Zapf Dingbats	ZDings	1989	1991	1.0	✏		U+270F	*pencil*
ARIB	ARIB	2007	2008	5.2	☔		U+2614	*umbrella with rain drops*
Japanese carriers	JCarrier	2007	2010	6.0	😎		U+1F60E	*smiling face with sunglasses*
Wingdings & Webdings	WDings	2010	2014	7.0			U+1F336	*hot pepper*

For a detailed view of when various source sets of emoji were added to Unicode, see emoji-versions-sources (the format is explained in Data Files). The correspondence to the original Japanese carrier symbols is in the UCD data file EmojiSources.txt. Data for the other sources is not in the UCD.

[Review Note: Add table showing when emoji were supported in various major products as background.]

A question often asked is how many emoji are in Unicode.This question does not have a simple answer, because there is no clear line separating which pictographic characters should and which should not be displayed with a typical emoji style. But roughly speaking, aside from the core set in Unicode Version 7.0 there are about 550 other characters that could also reasonably be displayed with typical emoji style (colored), such as U+1F46D TWO WOMEN HOLDING HANDS. There are also ways of representing emoji for national flags, adding about 240 others.

However, one of the goals of this document is to provide data for which Unicode characters would normally be considered to be emoji. *This data is currently just in draft form*, but according to that data, about 1,500 emoji are represented in Unicode 7.0. For more information, see Which Characters are Emoji.

1.1 Emoticons and Emoji

The term *emoticon* refers to a series of text characters (typically punctuation or symbols) that is meant to represent a facial expression or gesture (sometimes when viewed sideways), such as the following.

;-)

Emoticons predate Unicode and emoji, but were later adapted to include Unicode characters. The following examples use not only ASCII characters, but also U+203F (‿), U+FE35 (︵), U+25C9 (◉), and U+0CA0 (ಠ).

^‿^

◉︵◉

ಠ_ಠ

Often implementations allow emoticons to be used to input emoji. For example, the

emoticon ;-) can be mapped to 😉 in a chat window. The term *emoticon* is sometimes also used in a broader sense, to also include the emoji for facial expressions and gestures. That broad sense is used in the Unicode Block Name *Emoticons*, covering the code points from U+1F600 to U+1F64F.

1.2 Encoding Considerations

Unicode is the foundation for text in all modern software: it's how all mobile phones, desktops, and other computers represent all text of every language. People are using Unicode every time they type a key on their phone or desktop computer, and every time they look at a web page or text in an application. It is thus very important that the standard be stable, and that every character that goes into it be scrutinized carefully; thus there is a long development cycle for characters, with a formal process. For example, the 🕶 *dark sunglasses* character was first proposed years before it was released in Unicode 7.0.

To be considered for encoding, characters must normally be in widespread use as textual elements. The emoji and various symbols were added to Unicode because of their use as characters for text-messaging in a number of Japanese manufacturers' corporate standards, and other places, or in long-standing use in widely distributed fonts such as Wingdings and Webdings. In many cases, the characters were added for complete round-tripping to and from a source set, *not* because they were inherently of more importance than other characters. For example, the 🖰 *clamshell phone* character was included because it was in Wingdings and Webdings, not because it is more important than, say, a "skunk" character.

In some cases, a character was added to complete a set: for example, a 🏉 *rugby football* character was added to Unicode 6.0 to complement the 🏈 *american football* character (the ⚽ *soccer ball* had been added back in Unicode 5.2). Similarly, a mechanism was added that could be used to represent represent all country flags (those corresponding to a two-letter unicode_region_subtag), such as the 🇨🇦 *flag for Canada*, even though the Japanese carrier set only had 10 country flags.

People wanting to submit emoji or any other character for consideration for encoding should see the detailed instructions about how to submit character encoding proposals. It may be helpful to review the Unicode Forum or the Unicode Mail List, as well.

The historical documents used in the development of Unicode emoji from the Japanese carriers may be useful for comparison, since they show the original Japanese images and the first proposed reference glyphs. See the start of this section.

For more information about emoji, see the Unicode Emoji FAQ.

1.3 Goals

This document provides:

- design guidelines for improving interoperability across platforms and implementations
- data for which characters normally can be considered to be emoji

- data for which of those should be displayed by default with a text-style versus an emoji-style
- data for how to sort emoji characters more naturally
- data for useful categories for character-pickers for mobile and virtual keyboards
- data for useful annotations for searching emoji

It also provides background information about emoji, and discusses longer-term approaches to emoji.

[Review Note: This document does not discuss the issue of additional emoji characters after Unicode 7.0, whether for diversity or other purposes. However, the committee is considering additional variation selectors to indicate a preference among a small set of presentations for people emoji, such as male/female, or light/medium/dark skinned.]

As new Unicode characters are added or the “common practice” for emoji usage changes, the data and recommendations supplied by this document may change in accordance. Thus the recommendations and data will change across versions of this document.

This document does not discuss the issue of adding new emoji characters to Unicode after Unicode 7.0. Additions are being addressed by the Unicode Technical Committee. However, as any new characters are approved, this document should be updated as appropriate.

[Review Note: The data presented here is draft, and may change considerably before publication. Some the data presented here, such as collation or annotations, might end up in the Unicode CLDR project instead.]

2 Design Guidelines

Characters can have two kinds of presentation:

- an *emoji presentation,* with colorful and perhaps whimsical shapes, even animated
- a *text presentation,* such as black & white

More precisely, a text presentation is a simple foreground shape whose color which is determined by other information, such as setting a color on the text, while an emoji presentation determines the color(s) of the character, and is typically multicolored. In other words, when you change the text color in a word processor, a character with an emoji presentation will not change color.

Any Unicode character can be presented with text presentation, as in the Unicode charts. Both the name and the representative glyph in the Unicode chart should be taken into account when designing the apparance of the emoji, along with the images used by other vendors. The shape of the character can vary significantly. For example, here are just some of the possible images for U+1F36D LOLLIPOP, U+1F36E CUSTARD, U+1F36F HONEY POT, and U+1F370 SHORTCAKE:

While the shape of the character can vary significantly, designers should maintain the same “core” shape. For example, a U+1F36F HONEY POT encodes for a pictorial representation of a pot of honey, not for some semantic like "sweet". It would be unexpected to represent U+1F36F HONEY POT as a sugar cube, for example. Deviating too far from that core shape can cause interoperability problems: see accidentally-sending-friends-a-hairy-heart-emoji. Direction (whether a person or object faces to the right or left, up or down) should also be maintained where possible, because a change in direction can change the meaning: when sending 🐊🔫 “crocodile shot by police”, people expect any recipient to see the pistol pointing in the same direction as when they composed it. Similarly, the U+1F6B6 *pedestrian* should face to the left 🚶, not to the right.

General-purpose emoji for people and body parts should also not be given overly specific images: the general recommendation is to be as neutral as possible regarding race, ethnicity, and gender. Thus for the character U+1F64B *happy person raising one hand*, the recommendation is to use a neutral graphic like 🙋 instead of an overly-specific image like 🙋‍♀️. This includes the characters listed in the annotations chart under “human”. The representative glyph used in the charts, or images from other vendors may be misleading: for example, the construction worker 👷 may be male or female. For more information, see the Unicode Emoji FAQ.

Names of symbols such as BLACK MEDIUM SQUARE or WHITE MEDIUM SQUARE are not meant to indicate that the corresponding character must be presented in black or white, respectively; rather, the use of “black” and “white” in the names is generally just to contrast **filled** versus **outline** shapes, or a darker color fill versus a lighter color fill. Similarly, in other symbols such as the hands U+261A BLACK LEFT POINTING INDEX and U+261C WHITE LEFT POINTING INDEX, the words “white” and “black” also refer to outlined versus filled, and do not indicate skin color.

However, other color words in the name, such as YELLOW, typically provide a recommendation as to the emoji presentation, which should be followed to avoid interoperability problems.

[Review Note: Eventually we will need to update the core spec and FAQ to match the recommendations given here.]

Emoji characters may not always be displayed on a white background. To prevent the character from not being visually distinct from the background, they are often best given a faint, narrow contrasting border. Thus a Japanese flag would have a border so

that it would be visible on a white background, and a Swiss flag have a border so that it is visible on a red background.

Current practice is for emoji to have a square aspect ratio, deriving from their origin in Japanese. For interoperability, it is recommended that this practice be continued with current and future emoji.

Flag emoji characters are discussed in Annex B: Flags.

Combining marks may be applied to emoji, just like they can be applied to other characters. When that is done, the combination should take on an emoji presentation. For example, a 1️⃣ is represented as the sequence "1" plus an emoji variation selector plus U+20E3 COMBINING ENCLOSING KEYCAP. Systems are unlikely, however, to support arbitrary combining marks with arbitrary emoji. Aside from U+20E3, the following

- U+20E4 COMBINING ENCLOSING UPWARD POINTING TRIANGLE to indicate a warning
- U+20E0 COMBINING ENCLOSING CIRCLE BACKSLASH to indicate a prohibition.

For example, ped. ahead (pedestrian crossing ahead) can be represented as 🚶 + U+20E4, and no bike (no bicycles allowed) can be represented as 🚲 + U+20E0.

[Review Note: The recommended base characters would be associated with traffic signs and perhaps a few other characters. Should we have data listing those, so that implementations would know what to concentrate on?]

2.1 Gender

The following emoji have explicit gender, based on the name and explicit, intentional constrasts with other characters.

U+1F466 boy
U+1F467 girl
U+1F468 man
U+1F469 woman
U+1F474 older man
U+1F475 older woman
U+1F46B man and woman holding hands
U+1F46C two men holding hands
U+1F46D two women holding hands
U+1F6B9 mens symbol
U+1F6BA womens symbol

U+1F478 princess
U+1F46F woman with bunny ears
U+1F470 bride with veil
U+1F472 man with gua pi mao
U+1F473 man with turban
U+1F574 man in business suit levitating
U+1F385 father christmas

All others should be depicted in a gender-neutral way.

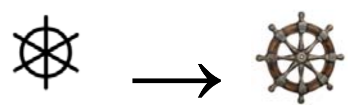
[Review Note: We may consider documenting for some of these that—*despite the name*—the depiction should be gender-neutral. Possibilities include: *father christmas, guardsman,* or *man in business suit levitating*.]

3 Which Characters are Emoji

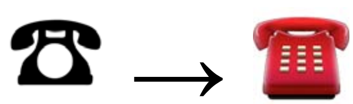
This document provides data in the Data Files for determining the set of characters which are expected to have an emoji presentation, either as a default or as a alternate presentation. This data was derived by starting with the characters that came from the original Japanese sets, plus those that major vendors have provided emoji fonts for. Characters that are similar to those in shape or design were then added. Often these characters are in the same Unicode blocks as the original set, but sometimes not.

While Unicode conformance allows any character to be given an emoji representation, characters that are not listed in the Data files should not normally be given an emoji presentation. For example, pictographic symbols such as keyboard symbols or math symbols (like ANGLE) that should never be treated as emoji. These are current recommendations: existing symbols can be added to this list over time.

This document takes a functional view as to the identification of emoji, which is that pictographs such as U+2388 HELM SYMBOL (introduced in Unicode 3.0) are categorized as emoji, since it is reasonable to give them either an emoji or text presentation, such as:



This follows the pattern set by characters such as U+260E BLACK TELEPHONE (introduced in Unicode 1.1), which can have either an emoji or text presentation, such as:



The data does not include *non*-pictographs, except for those in Unicode that are used to represent characters from emoji sources, such as:

 or 

Game pieces, such as the dominos (🀰 🀱 🀲 ... 🂒 🂓), are currently not included as emoji, with the exceptions of playing cards and mah jongg tiles. These are included because one of each set corresponds to an emoji character from one of the carrier sets.

[Review Note: We would like feedback on characters that should be added to this list in the Data Files, or removed from it. Removal would be warranted if the character is

really never suited for use in an emoji presentation.

Issue: should we also remove the playing cards and mah jongg tiles, except for the two from the carriers?

Issue: the following 7.0 characters appear to be redundant; should we also mark them as emoji? (The Symbola font can be installed if you can't see these.):

Issue: Other excluded punctuation and symbols can be reviewed to see whether or not they should be included, at other-labels.html.]

4 Presentation Style

Certain emoji have defined variation sequences, where an emoji character can be followed by one of two invisible variation selector

- U+FE0E for a text presentation
- U+FE0F for an emoji presentation

For more information on these selectors, see the file StandardizedVariants.html. Some systems may also provide this distinction with higher-level markup, rather than variation sequences.

Implementations should support both styles of presentation for the characters with variation sequences, if possible. Most of these characters were emoji that were unified with preexisting characters. Because people are now using emoji presentation for a broader set of characters, it is anticipated that more such variation sequences will be needed.

[Review Note: Wherever a character could reasonable be used with either presentation, variation sequences should be proposed for Unicode 8.0, scheduled for mid-2015.]

However, even where the variation selectors exist, it has not been clear for implementers 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 is important to have a shared sense among developers of when to use emoji presentation by default, so that there are fewer unexpected and "jarring" presentations. That is, to promote interoperability across platforms and applications, implementations need to know what the generally expected default presentation is.

That is, there has been no clear line for implementers between three categories of Unicode characters:

1. **emoji-default:** those expected to have an emoji presentation by default, but can also have a text presentation
2. **text-default:** those expected to have a text presentation by default, but could also have an emoji presentation
3. **text-only:** those that should only have a text presentation

The data files associated with this document provides data to distinguish between the first two categories: see the **Default** column of full-emoji-list. The data assignment is based upon current usage in browsers for Unicode 6.3 characters. For other characters, especially the new 7.0 characters, the assignment is based on that of the related emoji characters. For example, the “vulcan” hand 🖖 is marked as *emoji* because of the emoji styling currently given to other hands like ✋. The text-only characters are all those not listed in the data files.

In general, emoji characters are marked as text-default if they were in common use, and predated the use of emoji, and otherwise marked as emoji-default. For example, the negative squared A and B are text-default, while the negative squared AB is emoji-default. The reason is that A and B are part of a set of negative squared letters A-Z, while the AB was a new character. The default status may change over time, however, as a result in the change in usage.

The presentation of a given emoji character depends on the environment, whether or not there is an emoji or text variation selector, and the default presentation style (emoji vs text).

Enviroment	Presentation				Examples
	with Emoji VS	with Text VS	with no VS		
			text-default	emoji-default	
Text-Only	emoji	text	text	text	word processing document
Mixed	emoji	text	text	emoji	web pages
Presentation	emoji	text	emoji	emoji	texting, chats

[Review Note: We would like feedback on draft proposed default presentation in the Data Files: whether characters should have their defaults changed from emoji to text or vice versa.

TO_DO:

1. make most wing/webdings be #2.

]

5 Sorting

Neither the Unicode code point order, nor the standard Unicode Collation ordering (DUCET), are currently well suited for emoji, since they separate conceptually-related characters. For example, here is a selection of characters sorted by DUCET; to users this ordering appears quite random:

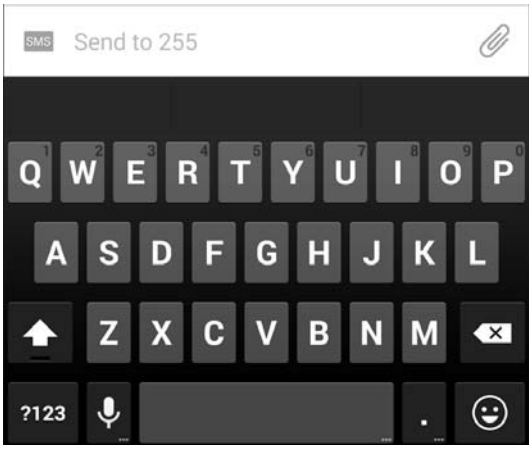
The emoji-ordering data file shows an ordering for emoji characters that groups them together in a more natural fashion.

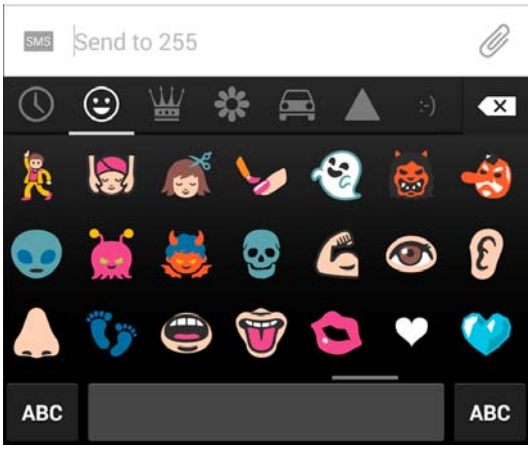
[Review Note: We would like feedback on the proposed ordering. The eventual

ordering would likely go into CLDR.]

6 Searching

Emoji are not typically typed on a keyboard. Instead, they are generally picked from a palette, or recognized via a dictionary. The mobile keyboards typically have a ☺ button to select a palette of emoji, such as in the left image below. Clicking on the ☺ button reveals a palette, as in the right image.





The palettes need to be organized in a meaningful way for users. 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. More advanced palettes will have long-press enabled, so that people can press-and-hold on an emoji and have a set of related emoji pop up. This allows for faster navitation, with less scrolling through the pallette.

Annotations for emoji characters are much more finely grained keywords. They can be used for searching characters, and are often easier than palettes for entering emoji characters. For example, when you type “hourglass” on your mobile phone, you could see and pick from either of the matching emoji characters ⏳ or ⌛. That is often much easier than scrolling through the palette and visually inspecting the screen. Input mechanisms may also map *emoticons* to emoji as keyboard shortcuts: typing :-) can result in 😄.

In some input systems, a word or phrase bracketed by colons is used to explicitly pick emoji characters. Thus typing in “I saw an *:ambulance:*” is converted to “I saw an 🚑”. For completeness, such systems might support all of the full Unicode names, such as *:first quarter moon with face:* for 🌛. Spaces within the phrase may be represented by _, as in the following:

“my *:alarm_clock:* didn’t work”

→

“my ⏰ didn’t work”.

However, in general the full Unicode names are not especially suitable for that sort of use; they were designed to be unique identifiers, and tend to be overly long or confusing.

Searching includes both searching for emoji characters in queries, and finding emoji characters in the target. These are most useful when they include the annotations as synonyms or hints. For example, when you search for ⛽ on yelp.com, you see matches for “gas station”. Conversely, searching for “gas pump” in a search engine could find pages containing ⛽. Similarly, searching for “gas pump” in an email program can bring up all the emails containing ⛽.

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”, a 😍 under “animal”, “cat”, and “heart”.

Annotations are language-specific: searching on yelp.de, you’d expect a search for ⛽ to result in matches for “Tankstelle”. Thus annotations need to be in multiple languages to be useful across languages. They should also include regional annotations within a given language, like “petrol station”, which you’d expect search for ⛽ to result in on yelp.co.uk. An English annotation cannot simply be translated into different languages, since different words may have different associations in different languages. The emoji 🌵 may be associated with Mexican or Southwestern restaurants in the US, but not be associated with them in, say, Greece. The scope of this document is limited to English annotations, but can provide an example for other languages.

There is one further kind of annotation, called a *TTS name*, for text-to-speech processing. For accessibility when reading text, it is useful to have a short, descriptive name for an emoji character. A Unicode character name can often serve as a basis for this, but its requirements for name uniqueness often ends up with names that are overly long, such as *black right-pointing double triangle with vertical bar* for ⏯. TTS names are also outside the current scope of this document.

TO_DO:

1. Divide into input and searching.

]

[Review Note: There is a suggestion for acronyms for each of the emoji. Feedback on this suggestion would be welcome.]

[Review Note: We would like feedback on changes to the annotations in the Data Files: additions, removals, or replacements. The eventual annotations would likely go into CLDR. One particular issue is whether or not to include forms of the same word: smile, smiles, smiling, smiled, smiley. The current policy is to only include a single form, assuming that any system using the annotations would handle related forms. However, the data has not been completely cleaned up to reflect that policy.]

7 Longer Term Solutions

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. Some examples of where this is done are:

- Captain America Skype Emoji

- Line Store
- Line Creators Market: Creation Guidelines
- Trello: Adding and removing stickers from cards

However, to be as effective and simple to use as emoji, a full solution requires significant infrastructure changes to allow simple, reliable input and transport of images (stickers) in texting, chat, mobile phones, email programs, virtual and mobile keyboards, and so on. Until that time, many implementations will need to use plain-text Unicode emoji instead.

For example, one necessary infrastructure change is to adapt mobile keyboards. Enabling embedded graphics would involve adding an additional custom mechanism for users to paste in their own graphics, such as a + sign to add an image to the palette above. This would prompt the user to paste or otherwise select a graphic, and add annotations for dictionary selection.

Once this is done, the user could then select those graphics in the same way as selecting the Unicode emoji. If users started adding many custom graphics, the mobile keyboard might even be enhanced to allow ordering or organization of those graphics so that they can be quickly accessed. The extra graphics would need to be disabled if the target of the mobile keyboard (such as an email header line) would only accept text.

Other features required to make embedded graphics work well include the ability of images to scale with font size, inclusion of embedded images in more transport protocols, switching services and applications to use protocols that do permit inclusion of embedded images (eg, MMS versus SMS for text messages). There will always, however, be places where embedded graphics can't be used—such as email headers, SMS messages, or filenames. There are also privacy aspects to implementations of embedded graphics: if the graphic itself is not packaged with the text, but instead is just a reference to an image on a server, then that server could track usage.

8 Media

There's been considerable media attention to emoji in 2014. There were some 6,000 articles on the emoji appearing in Unicode 7.0, according to Google News. Here are some examples of recent news about emoji (as of this writing):

Source	Title
Typographica	*Typeface Review: Apple Color Emoji*
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*
iDiversicons	*Emoticons for You... Representing an entire world of faces...*
CNET Japan	*Carriers unifying on Unicode Emoji* (machine-translated English version)
Vox	*Where Emoji come from*
Tom Scott	*Why Do Flag Emoji Count As Two Characters?*
The Wall Street Journal	*There's No Hot Dog Emoji, but New Characters Do Include a Hot Pepper*
NPR	*Why 140 Characters, When One Will Do? Tracing The Emoji Evolution*
Fast Company	*Where Do Emoji Come From?*
New Republic	*A Peek Inside the Non-Profit Consortium That Makes Emoji Possible*
Dissolve Footage	*Emoji Among Us: The Documentary*
Time	*Here Are Rules of Using Emoji You Didn't Know You Were Following*
Know Your Meme	*Emoticons*
New York Times	*The Emoji Have Won the Battle of Words*
Business Insider	*The Entire US Economy Depicted In Emoji*
Business Insider	*Scientists Describe Their Work In Emoji And The Results Are HilariousRead more: http://www.businessinsider.com /science-research-written-in-emoji-2014-10#ixzz3HNRL3ilt*
Spiegel	*Emoji-Siegeszug: Die Weltsprache unserer Zeit*
The New Inquiry	*The Conservatism of Emoji*
Technology Tell	*The 15 best emojis in Unicode 7.0*
CommitStrip	*Unicode 7 et ses nouveaux emoji*
Gizmodo	*My Day on the Emoji-Only Social Network, Translated (I Think)*
Oxford English Dictionary	*emoji, n.*
VentureBeat	*Microsoft's Bing beats Google to emoji search*

People have written online tools for seeing usage of emoji, such as *Emoji Tracker* and *Silicon Feelings*, and animations such as *emoji.zone*. It's also become popular to "translate" lyrics or sayings into the closest emoji, such as:

9 Data Files

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

TO_DO:

1. Move the descriptions of format to each file, and replace links to Data Files above to the specific data file.

]

The available files are:

File	Description
full-emoji-list	the main file: a list with images showing depictions from different sources, and the default status and annotations. For the column descriptions, see Full Emoji List.
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	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	an abbreviated list showing characters, not images. For checking browser/platform support.
emoji-style	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 Full Emoji List.*
emoji-labels	characters grouped by palette category. These are building blocks for palette categories, which would group some of these together.
emoji-annotations	characters grouped by annotation. *Also in column 7 of Full Emoji List.* The annotations are meant to be used in combination to winnow down the matches, so *:face moon:* would match the characters annotated with both "face" and with "moon".
emoji-ordering	draft ordering of emoji characters that groups like characters together. *Unlike the labels or annotations, each character only occurs once.*

other-labels	other general symbols and punctuation. That can be used to scan for other characters that might qualify for emoji presentation.
emoji-versions	a view of when different emoji were added to Unicode, by Unicode version.
emoji-versions-sources	a view of when different emoji were added to Unicode, and the sources. (See the **Version** information in Full Emoji List for the source description.) The sources indicate where a Unicode character corresponds to a character in the source. In many cases, the character had already been encoded well before the source was considered for other characters.

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

Typically, hovering over an image usually shows the code point and name, and clicking on the image goes to the respective row in the Full Emoji List. 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 browser. Your browser's zoom is also useful for examining the characters and images.

9.1 Full Emoji List

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

Column	Description
Count	A line count, for reference.
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 puts a link to that row in the address bar.
Browser	The plaintext character, showing whatever image would be native for the browser.
B&W*	The visual appearance of the codes, using the Unicode Chart font, plus PNGs for the flags.
Apple, Andr., Twit., Wind., GMail, SB, DCM, KDDI	Images from the respective sources for comparison. The GMail...KDDI are for comparison with images used before incorporation into Unicode. 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.

Name	The character name in lowercase (or an informative gloss, for the case of flags and keycaps).
Version	The version of Unicode in which the emoji was added (or will be, for Unicode 7.0). A superscript indicates the source of the character. Where a Unicode character corresponds to multiple sources, multiple superscripts will be present. The sources are:

z	ZDings	Zapf Dingbats
a	ARIB	
j	JCarrier	Japanese telephone carriers
w	WDings	Wingdings & Webdings
x	Other	other sources

Default	The draft proposed default presentation style. A * indicates that there are variation selectors (text and emoji) for the character.
Annotations	A rough-draft list of informative annotations. Clicking on a link goes to the respective row in the emoji-annotations.

Because the name and code point are already present, hovering or clicking on an image don't have the same effect as in other files. However, the alt values are still present for cut and paste into plaintext.

Annex A: Terminology

The goal for this annex is to collect the terminology used in connection to emoji for eventual incorporation into the Unicode Glossary.

Emoji - A colorful pictograph that can be used inline in text. Internally the representation is either (a) an image or (b) an encoded character. The term *emoji character* can be used for (b) where not clear from context.

Emoticon - (1) A series of text characters (typically punctuation or symbols) that is meant to represent a facial expression or gesture such as ;-) (2) a broader sense, also including emoji for facial expressions and gestures.

[Others TBD]

Annex B: Flags

There are 26 REGIONAL INDICATOR symbols that can be used in pairs to represent country flags. This mechanism was designed to be extensible, rather than be limited to just the 10 flags supported by the Japanese carriers.

Where flag emoji characters are supported, they should not just be limited to the 10 Japanese carrier flags. To avoid discriminating against other flags, they should instead be present for all of the valid country codes. More specifically, these are the BCP47 region subtags that are neither deprecated, nor private use, and nor macroregions (with the exception of the EU). This can be determined mechanically from data in CLDR. An overseas territory sometimes doesn't have its own flag, or only has flags for

subregions. In such cases, it may share the same flag as for the country.

Emoji are generally presented with a square aspect ratio, which presents a problem for flags. The flag for Qatar  is over 250% wider than tall; for Switzerland  it is square; for Nepal  it is over 20% taller than wide. To avoid a ransom-note effect, implementations may want to use a fixed ratio across all flags, such as 150%, with a blank band on the top and bottom. (The average width for flags is between 150% and 165%.)

For an open-source set of flag images (png and svg), see region-flags.

Acknowledgments

Mark Davis and Peter Edberg created the initial versions of this document, and maintain the text.

Thanks to Shervin Afshar, Norbert Lindenberg, Ken Lunde, Katsuhiko Momoi, Katrina Parrott, Markus Scherer, and Ken Whistler for feedback on this document, including earlier versions.

References

[Review Note: We'll flesh out the references later.]

[Unicode] The Unicode Standard
For the latest version, see:
http://unicode.org/versions/latest/

[UTR36] UTR #36: *Unicode Security Considerations*
http://unicode.org/reports/tr36/

[UTS39] UTS #39: *Unicode Security Mechanisms*
http://unicode.org/reports/tr39/

[Versions] Versions of the Unicode Standard
http://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

- First working draft based on Feb 2014 UTC discussions
- Added draft data files (as HTML for viewing)
- Updated text, changed files to use images for viewing across platforms.
- Updated as per May 2014 UTC discussion.
- Additions based on other feedback.
- Moved some background material from the introduction into Background; changes some lists into tables for ease of reading.

- Cleaned up the text based on feedback from the editorial committee.
- Incorporated some feedback from the August 2014 UTC; more to follow.
