Technical Reports

Working Draft Unicode Technical Report #51

UNICODE EMOJI

Version	1.0 (working draft)
Editors	Mark Davis (markdavis@google.com), Peter Edberg
Date	2014-07-24
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

This document provides information about emoji characters in Unicode, 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; useful annotations for searching emoji; and longer-term approaches to emoji.

It also presents recommendations for adding variation selectors for Unicode 8.0, and guidance for limiting glyphic variation to promote interoperability across platforms and implementations.

Status

*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 symbol characters) that can be presented in a
colorful form. 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 quite popular worldwide.

The word *emoji* comes from the Japanese:

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

Emoji became generally available in the early 2000s on Japanese cell phones.
There was an early proposal (2000) to encode DoCoMo emoji in Unicode. At that
time, it was unclear whether these characters would come into widespread use.

The emoji turned out to be quite popular, but each vendor developed different
(but partially overlapping) sets, and each cell 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).

To avoid the problem of multiple incompatible text encodings for emoji, and to enable interchange with Unicode-based systems, work began in the late 2000s to standardize the Japanese cell phone emoji in Unicode. A set of 722 characters was defined as the union of the emoji characters used by the various Japanese cell phone vendors; of these, 114 were mapped to characters already in Unicode, and the remaining 608 characters were added in Unicode 6.0, released in 2010. Several other emoji characters were added to Unicode at the same time.

Pictographs had been present in Unicode since 1993, but the the first Unicode characters explicitly intended as emoji were added for interoperability with the ARIB set in 2009 with Unicode 5.2. The largest group of emoji were then added in 2010 with Unicode 6.0. A few more pictographs were added in 2012 with version 6.1, and a large number were added with Unicode 7.0.

Here is a timeline of when some of the major sources of emoji were encoded in Unicode:

Source	Dev. Starts	Released	Unicode Version	Sample Character			
				B&W	Color	Code	Name
Zapf Dingbats	1989	1993	1.1	✏	✏️	U+270F	*pencil*
ARIB	2007	2008	5.2	☂	☔	U+2614	*umbrella with rain drops*
Japanese carriers	2007	2010	6.0	😎	😎	U+1F60E	*smiling face with sunglasses*
Wingdings & Webdings	2010	2014	7.0	🌶	🌶️	U+1F336	*hot pepper*

For a 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 a data file EmojiSources.txt.

1.1 Encoding Considerations

Unicode is the foundation for 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 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 to 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.

Some 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:

- emojidata.html - searchable doc, but the images for DoCoMo and SoftBank are currently blocked
- emojidata.pdf - full list, with all images

The following were earlier versions of the proposal for the carrier emoji.

- http://unicode.org/L2/L2007/07257-emoji-wd.html
- http://unicode.org/L2/L2009/09025r2-emoji.pdf
- http://unicode.org/L2/L2009/09026-emoji-proposed.pdf

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

1.2 Goals

This document provides information on:

- design guidelines for improving interoperability across platforms and implementations
- 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
- useful annotations for searching emoji
- longer-term approaches to emoji

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 supplied by successive versions of this document may change.

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.

[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.

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. Similarly, the original Unicode glyph for "pile of poo" is not a face, and does not have eyes. 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.

Flags should ideally be present for all of the BCP47 regions that are not deprecated, are not private use, and are not macroregions. This can be determined mechanically from data in CLDR. Flags for overseas territories 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 white band on the top and bottom. (The average width for flags is between 150% and 165%.) Flags are often best displayed with a faint border, otherwise the wrong impression of the shape is conveyed (especially for white sections): imagine the Qatar flag on a white background, or a Swiss flag on a red background.

3 Identification

This document provides a mechanism 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.

This document takes a functional view 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:
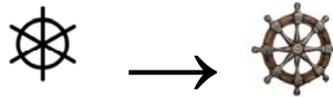
This follows the pattern set by characters such as U+260E BLACK TELEPHONE (introduced in Unicode 1.x), which can have either an emoji or text presentation, such as:
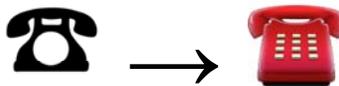
It does not add *non*-pictographs, even though some non-pictographs were incorporated into Unicode from emoji sources, such as:

 or

[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: 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: there seems to be little practical value to emoji dominos   …   , (since they are normally B&W), so they are currently excluded. 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.

[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.]

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. those expected to have an emoji presentation by default, but can also have a text presentation
2. those expected to have a text presentation by default, but could also have an emoji presentation
3. 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 🖐.

[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.

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 Data Files propose an ordering for emoji characters that groups them together in a more natural fashion.

[Review Note: We would like feedback on the proposed ordering in the Data Files. 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.
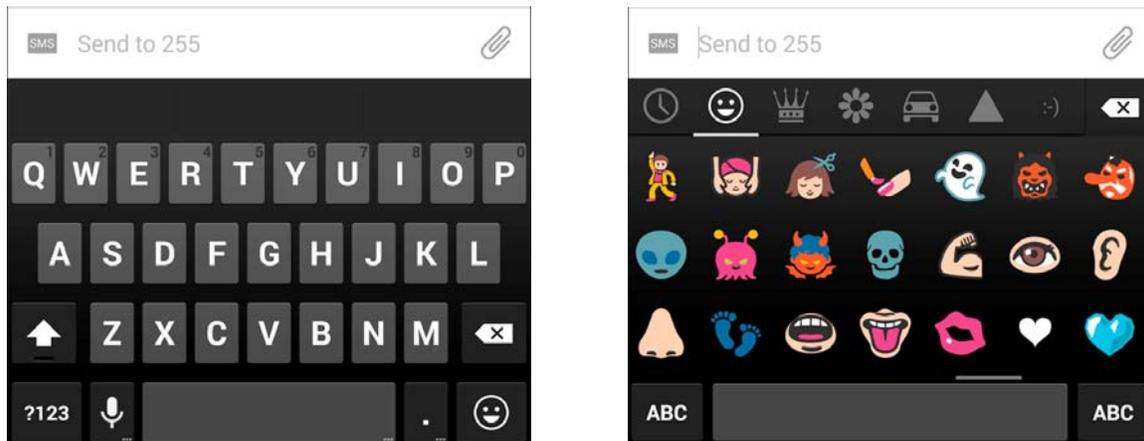
The palettes need to be organized in a meaningful way for users. They typically provide a small number of broad categories (5-10), such as People (anything associated with people), Nature, and so on. These categories typically have 100-200 emoji.

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 can support all of the full Unicode names, even where long, such as *:first quarter moon with face:* for 🌛. Spaces within the phrase may be represented by _, as in "my *:alarm_clock:* didn't work" → "my ⏰ didn't work".

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.

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.

<p align="center">;-)</p>

<p align="center">^‿^</p>

<p align="center">◉⁀◉</p>

<p align="center">ಠ_ಠ</p>

These examples use not only ASCII characters, but also U+203F (‿), U+FE35 (⁀), U+25C9 (◉), and U+0CA0 (ಠ). Emoticons may also be used as Emoji annotations, expecially for input. For example, the emoticon ;-) can be mapped to 😉 in a chat window. The term *emoticon* is sometimes used in a broader sense, to also include emoji for facial expressions and gestures.

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.

[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. An example of where this was done is Captain America Skype Emoji. However, this requires significant infrastructure changes to allow simple, reliable input and transport of images in texting, chat, mobile

phones, email programs, virtual and mobile keyboards, and so on. Until that time, implementations will typically 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*

[Review Note: These are useful to give context during development of this document. But we might remove them or move them elsewhere (such as the Unicode Emoji FAQ) if we think they'll go stale.]

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. 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, 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.
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 and 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 <u>emoji-annotations</u>.

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.

Acknowledgments

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

Thanks to 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.

liability is assumed for incidental and consequential damages in connection with or arising out of the use of the information or programs contained or accompanying this technical report. The Unicode Terms of Use apply.

Unicode and the Unicode logo are trademarks of Unicode, Inc., and are registered in some jurisdictions.