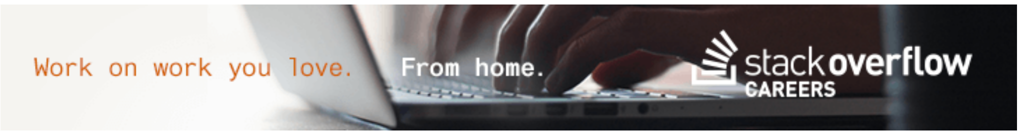
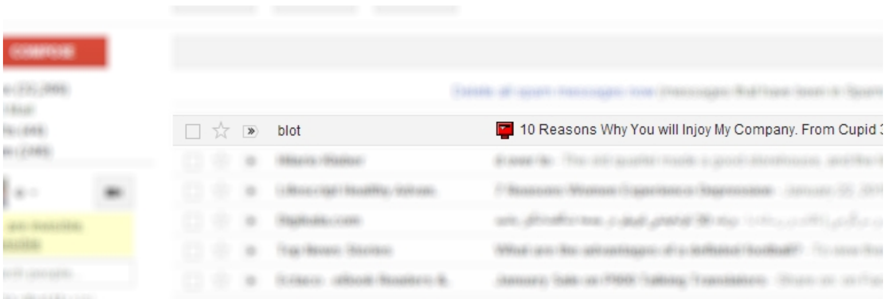
Animated icon in email subject

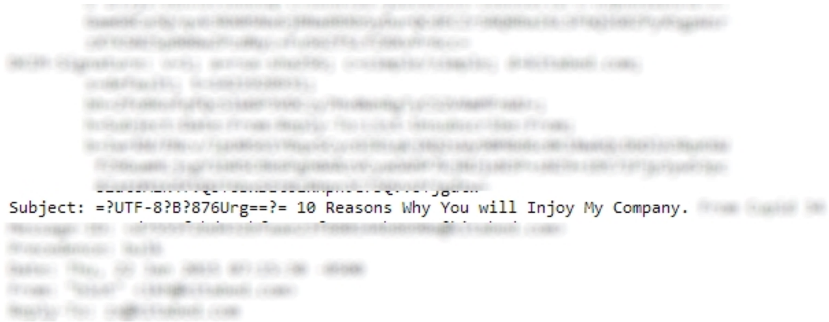


I know about **Data URIs** in which `base64` encoded data can be used inline such as images. Today I received an email actually a spam one in which there was an animated (gif) icon in its subject:



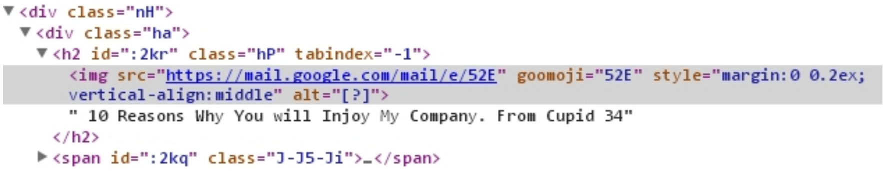
Here is the icon alone:

So the only thing did cross my mind was all about Data URIs and if Gmail allows some sort of emoticons to be inserted in subject. I saw the full detailed version of email and pointed to subject line at the below picture:



```
Subject: =?UTF-8?B?876Urg==?= 10 Reasons Why You will Injoy My Company.
```

So GIF comes from `=?UTF-8?B?876Urg==?=` encoded string that's similar to Data URI scheme however I couldn't get the picture out of it. Here is element HTML source:



```
▼<div class="nH">
  ▼<div class="ha">
    ▼<h2 id=":2kr" class="hP" tabindex="-1">
        <img src="https://mail.google.com/mail/e/52E" goomoji="52E" style="margin:0 0.2ex;
        vertical-align:middle" alt="[?]">
        " 10 Reasons Why You will Injoy My Company. From Cupid 34"
      </h2>
    ▶<span id=":2kq" class="J-J5-Ji">…</span>
```

Long story short, there are a lot of emoticons from `https://mail.google.com/mail/e/XXX` where `XXX` are hexadecimal values. They are documented nowhere or I couldn't find it. If that's about Data URI, so how is it possible to include them in Gmail's email subject? (I forwarded that email to a yahoo email account, seeing `[?]` instead of icon) and if it's not, then how that encoded string is parsed?

gmail base64 data-uri

edited Feb 4 '15 at 10:19

asked Jan 22 '15 at 17:47
revo
6,200 ●4 ●22 ●49

6 The real question is how do you block them?! – bambams Jul 9 '15 at 18:48

@bambams What do you mean? – revo Jul 9 '15 at 21:36

4 They are incredibly annoying and as you said they are only used by spammers. I'd rather they were just not shown by Gmail (it already seems to detect 99% as spam). – bambams Jul 17 '15 at 19:18

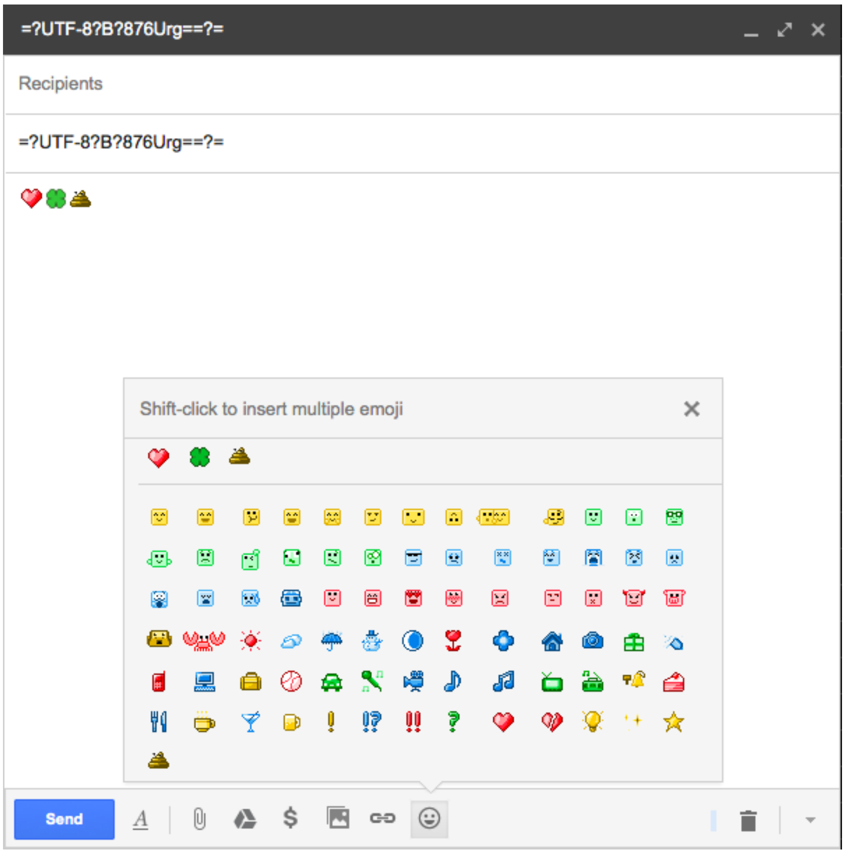
2 Answers

Short description:

They are referred to internally as `goomoji`, and they appear to be a non-standard UTF-8 extension. When Gmail encounters one of these characters, it is replaced by the corresponding icon. I wasn't able to find any documentation on them, but I was able to reverse engineer the format.

What are these icons?

Those icons are actually the icons that appear under the "Insert emoticons" panel.



=?UTF-8?B?876Urg==?=

Recipients

=?UTF-8?B?876Urg==?=

❤🍀💩

Shift-click to insert multiple emoji

While I don't see the `52E` icon in the list, there are several others that follow the same convention.

- `B0C` ❤
- `4F4` 💩

Note that there are also some icons whose names are prefixed, such as `gtalk.03C` 🍀. I was not able to determine if or how these icons can be used in this manner.

What is this Data URI thing?

It's not actually a Data URI, though it does share some similarities. It's actually a special syntax for encoding non-ASCII characters in email subjects, defined in RFC 2047. Basically, it works like this.

```
=?charset?encoding?data?=
```

So, in our example string, we have the following data.

```
=?UTF-8?B?876Urg==?=
```

- `charset` = `UTF-8`
- `encoding` = `B` (means base64)
- `data` = `876Urg==`

So, how does it work?

We know that somehow, `876Urg==` means the icon `52E`, but how?

If we base64 decode `876Urg==`, we get `0xf3be94ae`. This looks like the following in binary:

```
11110011 10111110 10010100 10101110
```

These bits are consistent with a 4-byte UTF-8 encoded character.

```
11110xxx 10xxxxxx 10xxxxxx 10xxxxxx
```

So the relevant bits are the following.:

```
     011   111110   010100   101110
```

Or when aligned:

```
00001111 11100101 00101110
```

In hexadecimal, these bytes are the following:

```
FE52E
```

As you can see, except for the `FE` prefix which is presumably to distinguished the `goomoji` icons from other UTF-8 characters, it matches the `52E` in the icon URL. Some testing proves that this holds true for other icons.

Sounds like a lot of work, is there a converter?:

This can of course be scripted. I created the following Python code for my testing. These functions can convert the base64 encoded string to and from the short hex string found in the URL. Note, this code is written for Python 3, and is not Python 2 compatible.

Conversion functions:

```
import base64

def goomoji_decode(code):
    #Base64 decode.
    binary = base64.b64decode(code)
    #UTF-8 decode.
    decoded = binary.decode('utf8')
    #Get the UTF-8 value.
    value = ord(decoded)
    #Hex encode, trim the 'FE' prefix, and uppercase.
    return format(value, 'x')[2:].upper()

def goomoji_encode(code):
    #Add the 'FE' prefix and decode.
    value = int('FE' + code, 16)
    #Convert to UTF-8 character.
    encoded = chr(value)
    #Encode UTF-8 to binary.
    binary = bytearray(encoded, 'utf8')
    #Base64 encode return end return a UTF-8 string.
    return base64.b64encode(binary).decode('utf-8')
```

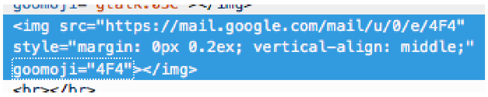
Examples:

```
print(goomoji_decode('876Urg=='))
print(goomoji_encode('52E'))
```

Output:

```
52E
876Urg==
```

And, of course, finding an icon's URL simply requires creating a new draft in Gmail, inserting the icon you want, and using your browser's DOM inspector.



```
goomoji="gtalk.03c"></img>
<img src="https://mail.google.com/mail/u/0/e/4F4"
style="margin: 0px 0.2ex; vertical-align: middle;"
goomoji="4F4"></img>
<hr></hr>
```

edited Dec 26 '15 at 8:28

answered Feb 4 '15 at 19:28
Alexander O'Mara
22.8k ● 7 ● 41 ● 60

4 That's an amazing complete answer. I don't have anything to say but I just wonder how did you do a reverse engineering on that!! Thank you Alexander. – revo Feb 5 '15 at 18:38

1 The assertion that `B` in the special syntax implies Base64 might have been a guess (the string at the end sort of looks like a Base64 encoded string, if you have seen those before); after which it's not that hard to notice that the four bytes follow one of the UTF-8 patterns for Unicode chars, esp because he's looking for Unicode. It's pretty cool detective work, all the same :) – sameers Sep 2 '15 at 3:45

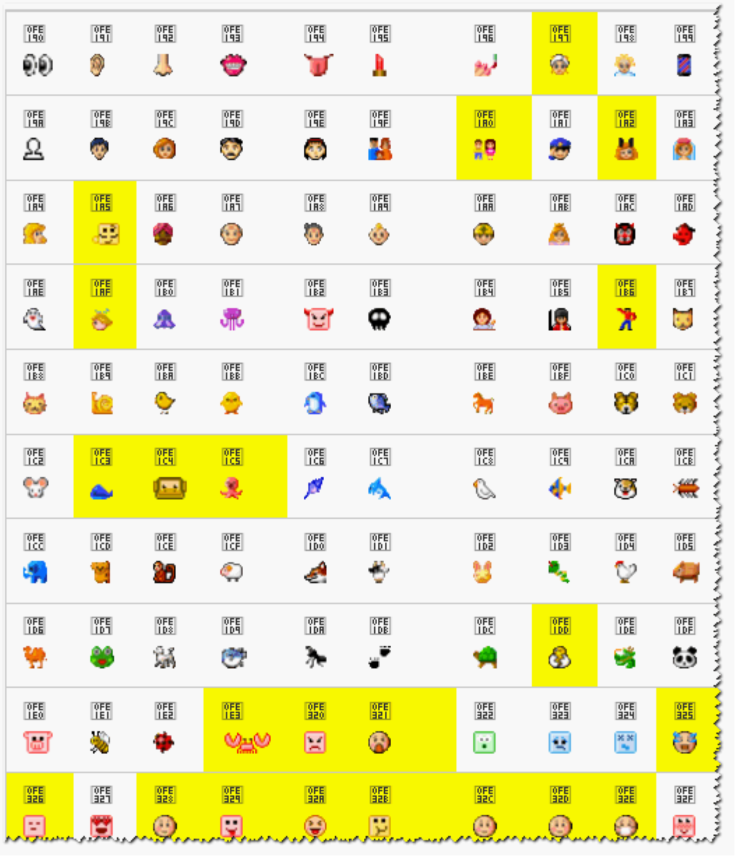
1 @sameers No need to guess about `B` -- it's defined in #4 – Jeremy Miller Dec 25 '15 at 22:24 ✎

It would be good to mention the RFC in the answer above, as a reference. – sameers Dec 26 '15 at 8:25

@JeremyMiller Thanks for tracking down the relevant RFC! I wasn't able to locate it when I was writing this answer. – Alexander O'Mara Dec 26 '15 at 8:28

If you use the correct hex code point (e.g. fe4f4 for 'pile of poo') and If it is correctly encoded within the subject line header, let it be base64 (see @AlexanderOMara) or quoted-printable (`=?utf-8?Q?=F3=BE=93=B4?=`), then Gmail will automatically parse and replace it with the corresponding emoji.

Here's a Gmail emoji list for copying and pasting into subject lines - or email bodies. Animated emojis, which will grab even more attention in the inbox, are placed on a yellow background:



0FE190	0FE191	0FE192	0FE193	0FE194	0FE195	0FE196	0FE197	0FE198	0FE199
0FE19A	0FE19B	0FE19C	0FE19D	0FE19E	0FE19F	0FE1A0	0FE1A1	0FE1A2	0FE1A3
0FE1A4	0FE1A5	0FE1A6	0FE1A7	0FE1A8	0FE1A9	0FE1AA	0FE1AB	0FE1AC	0FE1AD
0FE1AE	0FE1AF	0FE1B0	0FE1B1	0FE1B2	0FE1B3	0FE1B4	0FE1B5	0FE1B6	0FE1B7
0FE1B8	0FE1B9	0FE1BA	0FE1BB	0FE1BC	0FE1BD	0FE1BE	0FE1BF	0FE1C0	0FE1C1
0FE1C2	0FE1C3	0FE1C4	0FE1C5	0FE1C6	0FE1C7	0FE1C8	0FE1C9	0FE1CA	0FE1CB
0FE1CC	0FE1CD	0FE1CE	0FE1CF	0FE1D0	0FE1D1	0FE1D2	0FE1D3	0FE1D4	0FE1D5
0FE1D6	0FE1D7	0FE1D8	0FE1D9	0FE1DA	0FE1DB	0FE1DC	0FE1DD	0FE1DE	0FE1DF
0FE1E0	0FE1E1	0FE1E2	0FE1E3	0FE320	0FE321	0FE322	0FE323	0FE324	0FE325
0FE326	0FE327	0FE328	0FE329	0FE32A	0FE32B	0FE32C	0FE32D	0FE32E	0FE32F

answered May 20 '15 at 23:58

lukeA
15.9k ●3 ●9 ●21