




Problems concerning "U+1F471 WESTERN PERSON" in ISO-IEC 10646 FPDAM8

Karl Pentzlin – karl.pentzlin@europatastatur.de – 2010-04-16

(This paper accompanies the German comment T3 for ISO/IEC 10646:FPDAM8 [JTC1/SC2/WG2 N3790-DIN].)

The emoji WESTERN PERSON was proposed as Emoji e-1A4 as is shown in N3681 "Emoji symbols: Background data":

Internal ID	Symbol	Name & Annotations	DoCoMo	KDDI	SoftBank	Google
e-1A4	 U+1F46F proposed	WESTERN PERSON = woman with blond hair Temporary Notes: both male and female icons exist	[白人]	 #705 白人 U+EB13 SJIS-F3D4 JIS-7A56	 #290 #old444 西洋人 U+E515 SJIS-FBB5	U+FE1A4

According to this document, the KDDI set contains a head of a woman with blond hair but no corresponding man, while the SoftBank set contains a head of a man with blond hair but no corresponding woman.

Due to this, in N3723 "Summary of repertoire for FPDAM8 content of ISO/IEC 10646:2003", a single Unicode character is proposed:

1F471  WESTERN PERSON
= e-1A4

Enlarged, the icons from the original sources look like this:



There is no reason given for the unification into one character.

A unification of icons depicting a male vs. a female head or face is arguable for icons which denote a special profession or cultural role (like e-181 POLICE OFFICER or e-1B4 INFORMATION DESK PERSON).

However, icons not denoting such a role are more likely to be associated with a concrete person, which is either male or female.



Imagine a male KDDI customer sending a SMS:

I  MY NEW  WHOM I MET YESTERDAY!

to his buddy who is a SoftBank customer, who receives:

I  MY NEW  WHOM I MET YESTERDAY!

the recipient may be somehow irritated about the sexual preference about the sender, who probably did not intend to cause this. Even if the sender had added something like "HER NAME IS MARY", the message appears silly to the recipient.

On the other hand, it seems hard to find examples for text which are likely to contain Emoji (i.e. mainly short messages in a non-formal context), where the difference between  and  does not matter.

Thus, it is concluded that the unification is erroneous. Having a unified icon for "man with blond hair" and "woman with blond hair" is not only useless, it is also harmful to have such a symbol in Unicode: Such a character cannot be expressed by a single glyph that deploys its full semantic spectrum (i.e., in fact, it is not a character).

J

T

Thus, there are two scenarios:

1. The Emoji set as outlined in N3681 is an industry standard which has to be encoded in Unicode 1:1 despite to all errors and shortcomings, as this is absolutely necessary for the participating industry (e.g. for interoperability purposes).

Then, the unified character is to be included as compatibility character, being clearly marked so, and strongly discouraged for any actual use (except for two closed user groups, namely KDDI users and SoftBank users, employing different semantics within each of these groups).

However:

- No claim is known to the author of this paper that the Emoji list given in N3681 is a formally adopted industry standard at all, rather than an informal ad-hoc arrangement.
 - The paper intended as the final reference to the sources, N3728 "Emoji sources", does not refer to the Emoji list of N3681 and the e-xxx reference numbers given there at all. Instead, it references the separate original sources of the companies (DoCoMo, KDDI, and SoftBank). Consequently, the entry for the character as proposed in N3723 (FPDAM8) in N3728 is:
1F471 ; ; F3D4 ; FBB5
referencing nothing than the two different original characters as sources.
 - Several Japanese comments on other Emoji suggest that the Emoji table is not as fixed as N3681 suggests.
2. The Emoji are to be included into Unicode in a way that they are usable for everybody (as it is common for Unicode characters until now), even for KDDI users who want to communicate with SoftBank users.

Then, the two separate and different icons have to be mapped to two separate Unicode characters.

This can be done without problems. The mapping in N3728 "Emoji sources" simply is to be changed to:

1Fxxx ; ; F3D4 ;

1Fyyy ; ; ; FBB5

without changing anything at the general structure of N3728.

This way already was proposed by the German and Irish national body in N3607 "Towards an encoding of symbol characters used as emoji":

1F46F  MAN WITH BLOND HAIR
1F470  WOMAN WITH BLOND HAIR

The German national body requests in comment T4 (besides comment T3, which independently addresses the naming "WITH BLOND HAIR" instead of "WESTERN") to encode these two characters, instead of a single useless and harmful character "WESTERN PERSON".