

emoji and domain names

SAC095

<https://www.icann.org/en/system/files/files/sac-095-en.pdf>

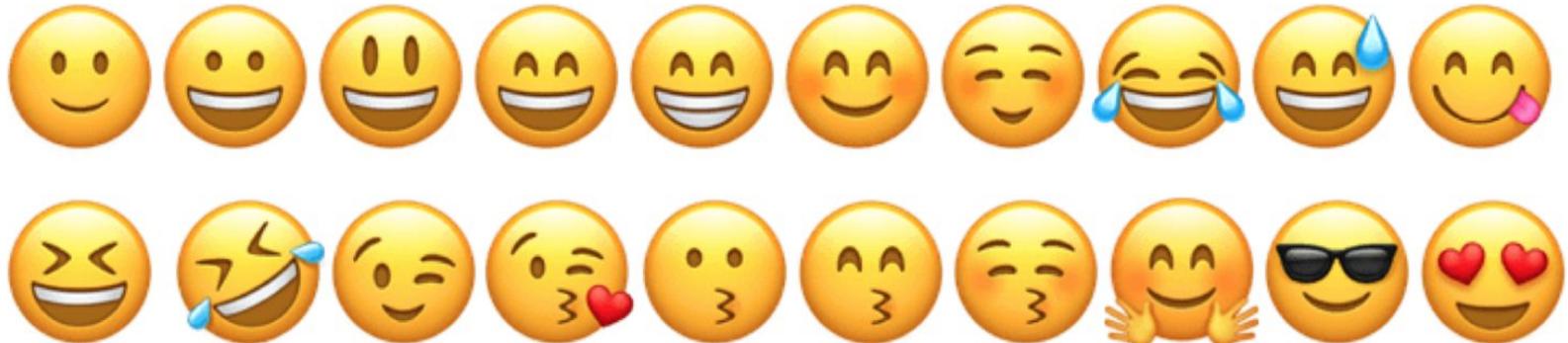
Patrik Fältström
SSAC

Emoji in IDNA

- IDNA is specified by the IETF in RFC 5890-5893
- The IDNA standard was designed to create, in conjunction with other methods to further narrow the set of acceptable labels, a set of identifiers that are, relative to a particular language community, acceptably reliable, easy to distinguish, and easy to input
- Characters in the Unicode Category “Symbol, Other” (So) were specifically not included; their derived property values are calculated as DISALLOWED in the IDNA standard
- Because emoji and other emoji-like symbol characters (e.g., 😊, 🍌) belong to the Unicode “So” category, they are disallowed by IDNA

Emoji Similarity

- Many emoji are visually similar and can be difficult to distinguish
- Distinguishability is not a design consideration when creating new emoji; ambiguity is acceptable
- Example: over twenty emoji with different code points are associated with “Smileys & People / face positive”
- See <http://unicode.org/emoji/charts/full-emoji-list.html>



Emoji Modifiers - Composition

- Some emoji can be “glued” together using a special zero width joiner (ZWJ) code point
- This allows multiple code points to be displayed as a single symbol



Emoji Modifiers - Modification

- In 2015 Unicode 8.0 introduced the ability to apply different colors to other emoji by appending one of five Fitzpatrick Skin Tone modifier code points to the code point for an anthropomorphic emoji



Universal Acceptance

- Universal Acceptance means that applications and systems must treat all Top Level Domains (TLDs), including new gTLDs and internationalized TLDs, in a consistent manner
- Specifically, they must accept, validate, store, process, and display all domain names unambiguously
- Additionally, there is a problem with accessibility, given that emoji are intrinsically visual constructs
- There is no agreed upon way to speak or enter an emoji, which makes them difficult to use with accessibility software

Findings

1. Emoji are disallowed by the IDNA standard
2. Emoji are not required by design, standard, or convention to be *visually uniform* or *visually distinguishable*
3. Emoji modifiers and “glue” arrangements allow for a potentially much larger set of composed multi-codepoint symbols
4. When two domain names are identical in appearance except for ordinary typographic style variations but have different underlying code points, they identify two different DNS domains
5. It is unrealistic to expect that just because a code point is included in Unicode, it should be used as part of a domain name

Recommendations

Recommendation 1: The SSAC recommends that the ICANN Board reject any TLD (root zone label) that includes emoji

Recommendation 2: The SSAC strongly discourages the registration of any domain name that includes emoji in any of its labels. The SSAC also advises registrants of domain names with emoji that such domains may not function consistently or may not be universally accessible as expected

After SAC095 was released: ICANN Board

- ICANN Board Meeting Thu, 2 Nov 2017
 - <https://features.icann.org/ssac-advisory-use-emoji-domain-names>
- Resolved (2017.11.02.09), the Board hereby directs that conformance to IDNA2008 and its successor will continue to be a necessary condition to determine valid IDN TLD labels
- Resolved (2017.11.02.10), the Board requests that the ccNSO and the GNSO engage with the SSAC to more fully understand the risks and consequences of using a domain name that includes emoji in any of its labels, and inform their respective communities about these risks
- Etc...

Personal comments

- Unicode Consortium runs an *Adopt-a-Character* program
 - <https://unicode.org/consortium/adopt-a-character.html>
- Early version of Unicode 11 is released for comments
 - Include a bunch of new emoji -- including “a lobster”
 - Global Domains International (GDI) “pre-sell” domain names including those
 - 161 single emoji sold to *emoji domain investors*
 - A .WS system glitch that allowed multiple registrations for the same emoji domains
- W3C I18n group is discussing CSS and modifications
 - CSS used for type family, style, and layout
 - But, also to sprinkle lines and dots over text (***text-decoration***)
 - This include adding modifiers, which makes the situation with emoji more interesting