

SAC095

SSAC Advisory on the Use of Emoji in Domain Names



An Advisory from the ICANN Security and Stability Advisory Committee (SSAC)
25 May 2017

Preface

This is an advisory to the ICANN Board, the ICANN community, and, more broadly, the Internet community from the ICANN Security and Stability Advisory Committee (SSAC) on the use of emoji in domain names.

The SSAC focuses on matters relating to the security and integrity of the Internet's naming and address allocation systems. This includes operational matters (e.g., pertaining to the correct and reliable operation of the root zone publication system), administrative matters (e.g., pertaining to address allocation and Internet number assignment), and registration matters (e.g., pertaining to registry and registrar services). SSAC engages in ongoing threat assessment and risk analysis of the Internet naming and address allocation services to assess where the principal threats to stability and security lie, and advises the ICANN community accordingly. The SSAC has no authority to regulate, enforce, or adjudicate. Those functions belong to other parties, and the advice offered here should be evaluated on its merits.

Table of Contents

1	Introduction	4
2	Emoji in Domain Names	4
2.1	Emoji in Internationalized Domain Names in Applications (IDNA)	4
2.2	Emoji Similarity	5
2.3	Emoji Modifiers	6
2.4	Universal Acceptance	7
3	Findings	8
4	Recommendations	9
5	Acknowledgments, Disclosures of Interest, Dissents, and Withdrawals	9
5.1	Acknowledgments	10
5.2	Disclosures of Interest	10
5.3	Dissents	10
5.4	Withdrawals	10

1 Introduction

Emoji¹ are “pictographs (pictorial symbols) that are typically presented in a colorful cartoon form and used inline in text.”² They are popular on smartphones, in chat, in email applications, and in social media, where they are part of a trend toward pictorial forms of communication to augment (or replace) text. Many single- and multiple-code point emoji have been included in the standard Unicode repertoire since version 6.0.³

Due to their popularity, there have been questions and discussions on the use of emoji in domain names.⁴ The Security and Stability Advisory Committee (SSAC) has studied this question and concluded that the use of emoji in any label of a domain name should be discouraged. In this advisory, the SSAC explains its reasoning.

2 Emoji in Domain Names

2.1 Emoji in Internationalized Domain Names in Applications (IDNA)

The standard for IDNA was developed by the Internet Engineering Task Force (IETF) and is documented in Requests for Comments (RFCs) 5890-5893. The IDNA standard⁵ 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.

In developing IDNA and considering security, stability, interoperability, and usability issues with an expanding character repertoire, the IETF chose to design a system with a limited number of *character types*. It used the “preferred name syntax” described in RFC 1034⁶ and extended the Letter-Digit-Hyphen (LDH) rules to allow for non-American Standard Code for Information Interchange (ASCII) letters, digits, and a few selected punctuations which were considered to be appropriate for Internet identifiers.

¹ The word *emoji* comes from the Japanese 絵 (e ≅ picture) 文 (mo ≅ writing) 字 (ji ≅ character). For additional background on emoji, please see section 1 of Unicode Technical Report 51 <<http://unicode.org/reports/tr51/>>

² See UTR#51, <<http://unicode.org/reports/tr51/>>

³ A number of symbols, formally *emoticons*, that are often used interchangeably with emoji have been in early versions of Unicode as well. For example, the frowning and smiling faces at U+2639..263B have been in Unicode since version 3.2.

⁴ Some country code Top Level Domains (ccTLDs), e.g., .WS (Samoa) and .LA (Laos), allow domain names with emoji to be registered at the second level, and several registrars accept labels with emoji for registrations in other TLDs. Since 2011, at least one browser displays any emoji it recognizes when they appear in domain names, although other browsers continue to show the ASCII A-labels instead.

⁵ See RFC 5892, <<https://www.rfc-editor.org/info/rfc5892>> for the current version (IDNA2008), and RFC 3490, <<https://www.rfc-editor.org/info/rfc3490>> for the previous obsolete version (IDNA2003).

⁶ See section 3.5 of RFC1034 on preferred name syntax.

SSAC Advisory on the Use of Emoji in Domain Names

Characters in the Unicode Category “Symbol, Other” (So) were specifically not included; their derived property values⁷ are calculated as DISALLOWED in the IDNA standard.⁸ Because emoji and other emoji-like symbol characters (e.g., ☺, 😊) belong to the Unicode “So” category, they are disallowed by IDNA.

Since most applications implement either the current version of IDNA (IDNA2008), the obsolete previous version (IDNA2003), or some locally synthesized combination of the two, the practical implication is that domain names with emoji will not be accepted or processed consistently by applications.

From an architectural perspective, if IDNs are to continue to be used as stable and secure identifiers, adherence to the original design goals of IDNA is needed.

This does not preclude or prohibit the use of emoji. Many other mechanisms exist⁹ for the usage of strings that are not intended to be used as identifiers, which may include symbols, special font adornments, and other characters that would be disallowed by IDNA. Because these strings are processed by mechanisms other than the Domain Name System (DNS), rules other than IDNA govern their composition and use. Where possible, these mechanisms should be used for emoji.

2.2 Emoji Similarity

Many emoji are visually similar and can be difficult to distinguish, especially when displayed in small fonts or by different applications, as no standard specifies exactly how they should be displayed. Distinguishability is not a design consideration when creating new emoji; ambiguity is acceptable.

One example of such ambiguity is the emoji categorized as “Smileys & People / face-positive”.¹⁰ Over twenty emoji with different code points are associated with this concept (see Figure 1), many of which look similar and can be virtually indistinguishable in smaller fonts.¹¹

⁷ In IDNA2008, each Unicode code point is assigned a value of either "PROTOCOL VALID", "CONTEXTUAL RULE REQUIRED", "DISALLOWED" or "UNASSIGNED". Sections 2 and 3 of RFC5892 contain the formula for calculating these values.

⁸ This is true for both the current (IDNA2008) and the previous obsolete (IDNA2003) versions of IDNA.

⁹ For example, clickable HyperText Markup Language (HTML) anchor text (e.g., the "I♥NY" in the HTML expression "I♥NY") would not be governed by IDNA2008, nor would a search term typed into a web search engine.

¹⁰ The original “smiling face” (U+263A) predates emoji; it and a few other “legacy” emoji are classified in the Unicode Category So block “miscellaneous symbols” rather than the So block “emoticons.”

¹¹ The Unicode Consortium publishes a chart <<http://unicode.org/emoji/charts/full-emoji-list.html>> that shows the way in which all of the Unicode emoji characters and sequences are displayed by popular browsing and messaging products.

SSAC Advisory on the Use of Emoji in Domain Names



Figure 1: Emoji symbols for Smileys & People / face-positive on Apple products

In ordinary conversational exchanges, these variations and ambiguities are perfectly acceptable. The sender of any “happy face” emoji is presumably trying to communicate happiness or a smile, and the receiver will understand this intention even if what they see is not identical to what the sender saw.¹²

However, when emoji are used in domain names, such ambiguities increase the risk of user confusion. It is generally difficult for people to figure out how to specify exactly what happy face they are trying to produce, and different systems represent the same emoji with different code points. The shape and color of emoji can change while a user is viewing them, and the user has no way of knowing whether what they are seeing is what the sender intended. As a result, the user is less likely to reach the intended resource and may instead be tricked by a phishing site or other intentional misrepresentation. It is likely to be even harder for the user to remember and type the exact emoji (or emoji sequence) that the registrant intended.

2.3 Emoji Modifiers

2.3.1 Composition

On top of the general confusability issues, some emoji can be “glued” together using a special zero width joiner (ZWJ) code point.¹³ This allows multiple code points to be displayed as a single symbol by systems that support the ZWJ (see Figure 2). Although this feature is beneficial for users looking for more creative ways to use emoji in other



Figure 2: An example of emoji with zero-width joiner (ZWJ) combining marks

¹² In practice, the interpretation of emoji by different viewers is much more complicated. Research has suggested that there is significant potential for miscommunication, both for individual emoji rendering and for different renderings across platforms; see, for example, “Sentiment of Emojis” at <<https://doi.org/10.1371/journal.pone.0144296>>; “Oh That’s What You Meant!: Reducing Emoji Misunderstanding” at <<http://www.sigchi.org/publications/toc/mobilehci-2016-adjunct>>; “Blissfully Happy’ or ‘Ready toFight’: Varying Interpretations of Emoji” at <<http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13167>>; and “Understanding Emoji Ambiguity in Context: The Role of Text in Emoji-Related Miscommunication” at <<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15703/14804>>

¹³ See <<http://www.unicode.org/reports/tr51/> - Emoji ZWJ Sequences>

contexts,¹⁴ it creates an additional confusability concern when emoji are used in identifiers.

The whole point of an identifier is to specify something unambiguously—*this* thing, as distinct from *all other* things. To a user, a single unmodified emoji might look exactly the same as its “glued together” counterpart, and systems that do not support emoji composition using a ZWJ will display the individual components of a “glued together” emoji as a sequence of separate emoji, with results that may visually be very different from what was intended. This is acceptable for interpersonal communication, particularly when it is augmented by shared context, but it is not acceptable for Internet identifiers, particularly DNS root labels that must be unambiguously resolved independent of any context.

2.3.2 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 points¹⁵ to the code point for an anthropomorphic emoji.¹⁶ The displayed results (see Figure 3) are highly sensitive to local conventions and interpretation.



Figure 3: Emoji constructed by appending skin tone modifiers

This presents an obvious confusability challenge to users who may find it difficult to detect these color differences, either because they do not perceive them (e.g., they are color-blind) or because the emoji are displayed in a way that makes them indistinguishable.

2.4 Universal Acceptance

Universal Acceptance for a truly multilingual Internet 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.¹⁷ Currently, it is already difficult to get people to accept the new labels that have appeared with IDNs and the new

¹⁴ See <<https://xkcd.com/1813/>>

¹⁵ See <<http://www.unicode.org/reports/tr51/> - Diversity>

¹⁶ Anthropomorphic emoji are defined here as emoji that are associated with human bodies and their activities. Section 2.2.2 “Sample Emoji Modifier Bases” of Unicode Technical Report 51 <<http://unicode.org/reports/tr51/>> lists some of the emoji that can receive skin tone modifiers.

¹⁷ See <<https://www.icann.org/resources/pages/universal-acceptance-2012-02-25-en>>

generic TLD (gTLD) program.¹⁸ Adding emoji to domain name labels will only make this problem worse and may reduce acceptability of non-ASCII labels due to their being perceived as simply too confusing or dangerous.

To protect their users, some applications will display the A-label (ASCII compatible encoding) of the domain name only, instead of the U-label (standard Unicode encoding).¹⁹ This further reduces the acceptability of domain names. As users find it hard to recognize and remember the corresponding A-label form of an IDN domain, many will become suspicious and decline to use or accept the IDN domain name.

Additionally, there is a problem with accessibility, given that emoji are intrinsically visual constructs. Existing IDNA characters represent text in natural languages, for which there are well developed accessibility techniques. However, there is no agreed upon way to speak or enter an emoji, which makes them difficult to use with accessibility software. Allowing supposedly global identifiers that cannot easily be rendered in a form that is accessible to visually impaired users is a significant problem.

3 Findings

Finding 1: Emoji are disallowed by the IDNA standard; domain names with emoji will not be accepted or processed consistently by applications.

Finding 2: Emoji are not required by design, standard, or convention to be visually *uniform* (one code point displayed the same way in all circumstances) or visually *distinguishable* (different code points displayed in ways that permit them to be disambiguated regardless of context). As a result, a user will be exposed to problems of confusability and accessibility. Different code points that are rendered the same or one code point that renders differently to different users will lead to inconsistent results depending on the display or rendering technology used.

Finding 3: Emoji modifiers and “glue” arrangements allow for a potentially much larger set of composed multi-codepoint symbols with even greater rendering variation and potential for ambiguous interpretation.

Finding 4: A fundamental property of the DNS is that it is an exact-match lookup service. For a given query, either there is a single name that matches or there is no match. When two domain names are identical in appearance except for ordinary typographic style variations (which, at present, have no equivalent for emoji), but have different underlying code points, they identify two different DNS domains.

Finding 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.

¹⁸ See <<https://uasg.tech/uasg-charter/>>

¹⁹ In this context, the terms A-label and U-label should be interpreted in accordance with the definitions specified in Section 2.3.2.1 of RFC 5890 <<https://tools.ietf.org/html/rfc5890-section-2.3.2.1>>.

SSAC Advisory on the Use of Emoji in Domain Names

While Unicode is used in the DNS, such usage should be considered secondary and for several reasons outlined in this report the repertoire must be limited.

Unicode, as an encoding system, is intended to accommodate a range of requirements for the encoding of natural language text (including printing). Natural language text has a number of flexibilities, including the assumption that readers will spot typographical and similar errors and be able to deduce what was intended from the context.

Identifiers such as domain names do not generally have such associated context. This is particularly true for TLD labels, where labels in the root zone do not have any linguistic context.²⁰

Finally, domain names are used by end users. In ordinary circumstances, they need to be constructed in such a way as to allow easy and accurate transcription by the end user from one context to another.

These factors constrain the number and the classes of Unicode characters suitable to be used as part of a domain name.

4 Recommendations

Recommendation 1: Because the risks identified in this Advisory cannot be adequately mitigated without significant changes to Unicode or IDNA (or both), the SSAC recommends that the ICANN Board reject any TLD (root zone label) that includes emoji.

Recommendation 2: Because the risks identified in this Advisory cannot be adequately mitigated without significant changes to Unicode or IDNA (or both), 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.

5 Acknowledgments, Disclosures of Interest, Dissents, and Withdrawals

In the interest of transparency, these sections provide the reader with information about four aspects of the SSAC process. The Acknowledgments section lists the SSAC members, outside experts, and ICANN staff who contributed directly to this particular document. The Disclosures of Interest section points to the biographies of all SSAC members, which disclose any interests that might represent a conflict—real, apparent, or potential—with a member’s participation in the preparation of this advisory. The Dissents section provides a place for individual members to describe any disagreement that they may have with the content of this document or the process for preparing it. The Withdrawals section identifies individual members who have recused themselves from discussion of the topic with which this advisory is concerned. Except for members listed

²⁰ See section 2.1 of SAC060 at <<https://www.icann.org/en/system/files/files/sac-060-en.pdf>>

SSAC Advisory on the Use of Emoji in Domain Names

in the Dissents and Withdrawals sections, this document has the consensus approval of all of the members of SSAC.

5.1 Acknowledgments

The committee wishes to thank the following SSAC members and external experts for their time, contributions, and review in producing this Advisory.

SSAC members

Jaap Akkerhuis
Lyman Chapin
Patrik Fältström
James Galvin
Julie Hammer
Geoff Huston
John Levine
Tara Whalen
Suzanne Woolf
Ram Mohan

Invited Experts

John Klensin
Pete Resnick

ICANN staff

Patrik Jones
Andrew McConachie
Kathy Schnitt
Steve Sheng (editor)

5.2 Disclosures of Interest

SSAC member biographical information and Disclosures of Interest are available at: <https://www.icann.org/resources/pages/ssac-biographies-2017-02-16-en>.

5.3 Dissents

There were no dissents.

5.4 Withdrawals

There were no withdrawals.