

SAC095

SSAC Consultivo sobre o Uso de *emoji* em Nomes de Domínio



Comunicado do Comitê Consultivo da ICANN para Segurança e Estabilidade (SSAC),
25 de maio de 2017

Prefácio

Este é um comunicado do Comitê Consultivo para Segurança e Estabilidade (SSAC) à Diretoria da ICANN, à comunidade da ICANN e, mais amplamente, à comunidade da Internet sobre o uso de *emojis* em nomes de domínio.

O SSAC concentra-se em questões relacionadas a segurança e integridade dos sistemas de alocação de nomes e endereços da Internet. Isso inclui questões operacionais (por exemplo, as relativas à operação correta e confiável do sistema de publicação da zona raiz de nomes), questões administrativas (por exemplo, as referentes à alocação de endereços e atribuição de números da Internet) e questões de registro (por exemplo, as referentes a operadores de registro e registradores). O SSAC envolve-se na avaliação contínua de ameaças e na análise de risco dos serviços de alocação de nomes e endereços da Internet, para identificar onde estão as principais ameaças à estabilidade e segurança, e assessorar a comunidade da ICANN. O SSAC não tem autoridade para regular, fazer cumprir ou julgar. As recomendações oferecidas são de mérito.

Sumário

Prefácio	2
1 Introdução	4
2 <i>Emoji</i> em Nomes de Domínio.....	4
2.1 <i>Emoji</i> em Nomes de Domínio Internacionalizados em Aplicações (IDNA)	4
2.2 Similaridade de <i>Emojis</i>	5
2.3 Modificadores de <i>Emoji</i>	6
2.4 Aceitação Universal.....	7
3 Conclusões.....	8
4 Recomendações.....	9
5 Agradecimentos, Divulgações de Interesse, Divergências e Retiradas....	9
5.1 Agradecimentos.....	10
5.2 Divulgações de Interesse	10
5.3 Divergências.....	10
5.4 Abstenções.....	11

1 Introdução

Os *emojis*¹ são pictogramas (símbolos pictóricos) tipicamente apresentados em forma de desenhos coloridos e utilizados em conformidade com o texto.² São populares em *smartphones*, *chats*, aplicações de e-mail e na mídia social, onde são parte de uma tendência de formas pictóricas de comunicação para realçar (ou substituir) o texto. Muitos *emojis* de ponto de código único e múltiplo foram incluídos no repertório padrão do Unicode desde a versão 6.0.³

Devido à sua popularidade, tem havido questões e discussões sobre o uso de *emojis* em nomes de domínio.⁴ O Comitê Consultivo para Segurança e Estabilidade (SSAC) estudou essa questão e concluiu que o uso de *emoji* em qualquer rótulo de um nome de domínio deve ser desencorajado. Neste comunicado, o SSAC explica seus motivos.

2 *Emoji* em Nomes de Domínio

2.1 *Emoji* em Nomes de Domínio Internacionalizados em Aplicações (IDNA)

O padrão para IDNA foi desenvolvido pela Força-Tarefa de Engenharia da Internet (IETF) e está documentado nas Solicitações de Comentários (RFCs) 5890 e 5893. O padrão IDNA⁵ foi projetado para criar, em conjunção a outros métodos para restringir ainda mais o conjunto de rótulos aceitáveis, um conjunto de identificadores que são, em relação a uma determinada comunidade linguística, aceitavelmente confiáveis, fáceis de distinguir e fáceis de inserir.

Ao desenvolver o IDNA e considerar questões de segurança, estabilidade, interoperabilidade e usabilidade com um repertório de caracteres em expansão, o IETF optou por definir um sistema com um número limitado de *tipos de caracteres*. Utilizou a “sintaxe de nome preferencial” descrita na RFC 1034⁶ e ampliou as regras de Letra-Dígito-Hífen (LDH) para permitir letras, dígitos e algumas pontuações selecionadas não incluídas no Código Padrão Americano para Intercâmbio de Informações (ASCII), mas que foram consideradas apropriadas para identificadores da Internet.

¹ A palavra *emoji* vem do japonês 絵 (e ≅ figura) 文 (mo ≅ escrita) 字 (ji ≅ caractere). Para informações adicionais sobre *emoji*, consulte a seção 1 do Relatório Técnico Unicode 51. <<http://unicode.org/reports/tr51/>>

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

³ Um número de símbolos, formalmente *emoticons*, muitas vezes utilizados de forma intercambiável com *emoji*, também apareceram nas primeiras versões do Unicode. Por exemplo, os rostos carrancudos e sorridentes em U+2639..263B estão em Unicode desde a versão 3.2.

⁴ Alguns domínios de topo de código de país (ccTLDs), por exemplo, .WS (Samoa) e .LA (Laos), permitem que nomes de domínio com *emoji* sejam registrados no segundo nível, e vários registradores aceitam rótulos com *emoji* para registros em outros TLDs. Desde 2011, pelo menos um navegador exibe quaisquer *emoji* que reconheça quando estes aparecem em nomes de domínio, embora outros navegadores continuem, em vez disso, exibindo os rótulos-A ASCII.

⁵ Ver RFC 5892, <<https://www.rfc-editor.org/info/rfc5892>> para a versão atual (IDNA2008), e RFC 3490, <<https://www.rfc-editor.org/info/rfc3490>> para a versão obsoleta anterior (IDNA2003).

⁶ Ver seção 3.5 da RFC 1034 sobre sintaxe de nome preferencial.

Caracteres na Categoria Unicode “Símbolo, Outro” (So) não foram especificamente incluídos; seus valores de propriedade derivados⁷ são calculados como NÃO PERMITIDO no padrão IDNA.⁸ Por pertencerem à categoria Unicode “So” (por exemplo, ☺, 😊), *emoji* e outros caracteres simbólicos do tipo *emoji* não são permitidos pelo IDNA.

Como a maioria das aplicações implementa a versão atual do IDNA (IDNA2008) e a versão anterior obsoleta (IDNA2003) ou alguma combinação local sintetizada dos dois, a implicação prática é que os nomes de domínio com *emoji* não serão aceitos ou processados consistentemente pelas aplicações.

Do ponto de vista da arquitetura, se os IDNs continuarem a ser utilizados como identificadores estáveis e seguros, a aderência às metas originais de projeto do IDNA será necessária.

Isso não impede ou proíbe o uso de *emoji*. Existem muitos outros mecanismos⁹ para a utilização de sequências que não devem ser empregadas como identificadores e que podem incluir símbolos, adornos especiais de fontes e outros caracteres que não seriam permitidos pelo IDNA. Como essas cadeias de caracteres são processadas por mecanismos diferentes do DNS, outras regras além do IDNA controlam sua composição e uso. Quando possível, esses mecanismos devem ser utilizados para o *emoji*.

2.2 Similaridade de *Emojis*

Muitos *emojis* são visualmente semelhantes e podem ser difíceis de distinguir, especialmente quando exibidos em fontes pequenas ou por diferentes aplicações, já que nenhum padrão especifica exatamente como devem ser exibidos. Distinguilidade não é uma consideração de projeto ao criar novos *emojis*; ambiguidade é aceitável em *emojis*.

Um exemplo dessa ambiguidade é o *emoji* caracterizado como “Sorrisos e Pessoas com face feliz”.¹⁰ Mais de 20 *emojis* com diferentes pontos de código são associados a esse conceito (ver figura 1), muitos dos quais parecem semelhantes e podem ser virtualmente indistinguíveis em fontes menores.¹¹

⁷ No IDNA 2008, a cada ponto de código Unicode é atribuído um valor de “PROTOCOLO VÁLIDO”, “REGRA CONTEXTUAL SOLICITADA”, “NÃO PERMITIDO” OU “NÃO ATRIBUÍDO”. As seções 2 e 3 da RFC5892 contêm a fórmula para calcular esses valores.

⁸ Isso é verdadeiro para ambas as versões de IDNA, tanto a atual (IDNA2008) quanto a anterior e obsoleta (IDNA2003).

⁹ Por exemplo, o texto âncora clicável Linguagem de Marcação de Hipertexto (HTML) (por exemplo, o “I♥NY” na expressão HTML “<a href=“<https://www.iloveny.example>”>I♥NY”) e um termo de busca digitado em um mecanismo de busca da Web não seriam governados pelo IDNA2008.

¹⁰ A “face sorridente” original (U+263A) antecede o *emoji*; tanto esse quanto alguns outros *emoji* “herdados” são classificados no bloco da Categoria So do Unicode “símbolos diversos” em lugar do bloco So de “emoticons”.

¹¹ O Consórcio Unicode publica uma tabela <<http://unicode.org/emoji/charts/full-emoji-list.html>> que demonstra como todos os caracteres e sequências *emoji* são exibidos por produtos populares de mensagem e navegação.

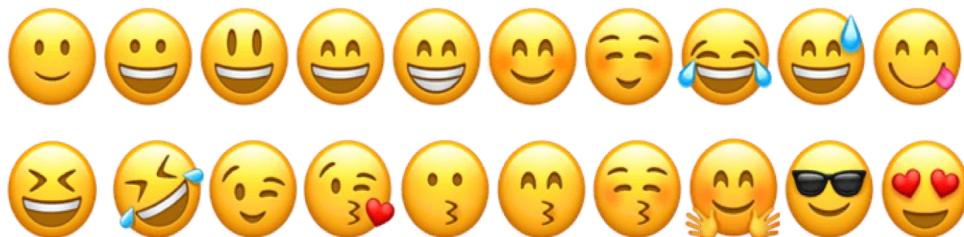


Figura 1: Símbolos *Emoji* para Smileys e Pessoas / face-positiva em produtos Apple

Em trocas conversacionais comuns, essas variações e ambiguidades são perfeitamente aceitáveis. O remetente de qualquer *emoji* “face feliz” está presumivelmente tentando comunicar felicidade ou um sorriso e o receptor entenderá essa intenção ainda que o que ele vê não seja idêntico ao que o remetente viu.¹²

No entanto, quando os *emojis* são utilizados em nomes de domínio, tais ambiguidades aumentam o risco de confusão dos usuários. Geralmente é difícil para as pessoas descobrirem qual face feliz estão tentando produzir, e sistemas diferentes representam o mesmo *emoji* com pontos de código diferentes. O formato e a cor dos *emojis* podem mudar enquanto um usuário os está vendo e os usuários não têm como saber se o que estão vendo é o que o remetente pretendia. Como resultado, é menos provável que o usuário alcance o recurso pretendido e pode, em vez disso, ser enganado por um *phishing* em um sítio ou outra deturpação intencional. É provável que seja ainda mais difícil para o usuário lembrar-se e digitar o *emoji* exato (ou a sequência *emoji*) que o registrador pretendia.

2.3 Modificadores de *Emoji*

2.3.1 Composição

Além dos problemas gerais de confundibilidade, alguns *emojis* podem ser “colados” por meio do uso de um ponto de código *zero width joiner* (ZWJ) especial¹³. Isso permite que vários pontos de código sejam exibidos como um único símbolo por sistemas que suportam o ZWJ (ver figura 2). Embora esse recurso seja benéfico para usuários em busca de maneiras mais criativas de utilização de *emoji* em outros contextos,¹⁴ cria uma preocupação adicional quanto à confundibilidade quando os *emojis* são utilizados em identificadores.

¹² ¹² Na prática, a interpretação dos *emojis* por visualizadores diferentes é muito mais complicada. A pesquisa sugere que há um potencial significativo de falhas de comunicação, tanto para renderização individual de *emoji* quanto para renderizações diferentes através das plataformas; ver, por exemplo, “Sentiment of Emojis” em <<https://doi.org/10.1371/journal.pone.0144296>>; “Oh That’s What You Meant!: Reducing Emoji Misunderstanding” em <<http://www.sigchi.org/publications/toc/mobilehci-2016-adjunct>>; “Blissfully Happy’ ou ‘Ready to Fight’: Varying Interpretations of Emoji” em <<http://www.aaii.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13167>>; e “Understanding Emoji Ambiguity in Context: The Role of Text in Emoji-Related Miscommunication” em <<https://www.aaii.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15703/14804>>

¹³ Ver <<http://www.unicode.org/reports/tr51/> - Emoji_ZWJ_Sequences>

¹⁴ Ver <<https://xkcd.com/1813/>>



Figura 2: Um exemplo de *emoji* com marcações combinadas de *zero width joiner* (ZWJ)

O principal objetivo de um identificador é especificar algo sem ambiguidade - *essa* coisa, distinta de todas as outras coisas. Para um usuário, um único *emoji* não modificado pode parecer exatamente igual ao seu correspondente "colado", e sistemas que não suportam composições de *emoji* utilizando um ZWJ exibirão os componentes individuais de um *emoji* "colado" como sequência de *emojis* separados, com resultados que podem ser visualmente muito diferentes do que se pretendia. Isso é aceitável para comunicação interpessoal, particularmente quando é ampliado pelo contexto compartilhado, mas não é aceitável para identificadores da Internet, particularmente rótulos de raiz do DNS que devem ser resolvidos sem ambiguidade independentemente de qualquer contexto.

2.3.2 Modificação

Em 2015 o Unicode 8.0 introduziu a habilidade de aplicar cores diferentes a outro *emoji* anexando um dos cinco pontos de código modificadores Fitzpatrick Skin Tone¹⁵ ao ponto de código para um *emoji* antropomórfico.¹⁶ Os resultados exibidos (ver figura 3) são altamente sensíveis a convenções e interpretações locais.



Figura 3: *Emoji* construído pela adição de modificadores de tom de pele

Isso apresenta um desafio óbvio de confundibilidade para os usuários, que podem considerar difícil detectar essas diferenças de cores, seja porque não as percebem (por exemplo, são daltônicos) ou porque os *emojis* são exibidos de uma maneira que os torna indistinguíveis.

2.4 Aceitação Universal

A Aceitação Universal para uma Internet realmente multilíngue significa que as aplicações e os sistemas devem tratar de maneira consistente todos os Domínios de Nível de Topo (TLDs), incluindo novos gTLDs e TLDs internacionalizados. Especificamente, devem aceitar, validar, armazenar, processar e exibir sem ambiguidade todos os nomes de

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

¹⁶ *Emojis* antropomórficos são definidos aqui como *emojis* associados a corpos humanos e suas atividades. A seção 2.2.2 "Sample Emoji Modifier Bases" do Relatório Técnico Unicode 51 <<http://unicode.org/reports/tr51/>> lista alguns *emojis* que podem receber modificadores de tom de pele.

domínio.¹⁷ Atualmente já é difícil levar as pessoas a aceitarem os novos rótulos que apareceram com os IDNs e o novo programa de TLDs genéricos (gTLD).¹⁸ A adição de *emoji* a rótulos de nomes de domínio só irá piorar esse problema e poderá reduzir a aceitabilidade de rótulos não ASCII por serem considerados simplesmente muito confusos ou perigosos.

Para proteger seus usuários, algumas aplicações irão exibir o rótulo-A (codificação compatível com ASCII) apenas do nome de domínio, em vez do rótulo-U (codificação padrão Unicode).¹⁹ Isso reduz ainda mais a aceitabilidade dos nomes de domínio. Como os usuários acham difícil reconhecer e lembrar-se do formulário do rótulo-A correspondente de um domínio IDN, muitos ficarão desconfiados e irão recusar-se a utilizar ou aceitar o nome de domínio IDN.

Além disso, há um problema com a acessibilidade, dado que os *emojis* são construções intrinsecamente visuais. Os caracteres IDNA existentes representam o texto em linguagens naturais, para as quais existem técnicas de acessibilidade bem desenvolvidas. No entanto, não há maneira acordada de falar ou inserir um *emoji*, o que dificulta o uso com software de acessibilidade. Permitir identificadores supostamente globais não passíveis de ser facilmente renderizados em um formato acessível a usuários com deficiência visual é um problema significativo.

3 Conclusões

Conclusão 1: Os *emojis* não são permitidos pelo padrão IDNA; nomes de domínio com *emoji* não serão aceitos ou processados consistentemente pelas aplicações.

Conclusão 2: Não é exigido em termos de desenho, padrão ou por convenção que os *emojis* sejam visualmente *uniformes* (um ponto de código exibido da mesma forma em todas as circunstâncias) ou visualmente *distinguível* (diferentes pontos de código exibidos de forma a permitir que sejam desambiguados, independentemente do contexto). Como resultado, um usuário será exposto a problemas de confundibilidade e acessibilidade. Diferentes pontos de código que são processados da mesma forma ou um ponto de código que é processado de maneira diferente para diferentes usuários levarão a resultados inconsistentes, dependendo da exibição ou da tecnologia de renderização utilizada.

Conclusão 3: Modificadores de *emoji* e arranjos de "cola" permitem um conjunto potencialmente maior de símbolos com múltiplos pontos de código compostos, com variação de renderização ainda maior e potencial para interpretações ambíguas.

Conclusão 4: Uma propriedade fundamental do DNS é que se trata de um serviço de pesquisa de correspondência exata. Há um único nome que corresponde a uma determinada consulta ou não há correspondência. Quando dois nomes de domínio são idênticos na aparência, exceto para variações de estilo tipográficas comuns (que, no

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

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

¹⁹ Nesse contexto, os termos rótulo-A e rótulo-U podem ser interpretados de acordo com as definições especificadas na Seção 2.3.2.1 da RFC 5890 <<https://tools.ietf.org/html/rfc5890-section-2.3.2.1>>.

momento, não têm equivalente para *emoji*), mas têm pontos de código subjacentes diferentes, eles identificam dois diferentes domínios DNS.

Conclusão 5: Não é realista esperar que apenas porque um código de ponto está incluído no Unicode deve ser utilizado como parte de um nome de domínio.

Enquanto o Unicode é utilizado no DNS, tal utilização deve ser considerada secundária e por várias razões destacadas neste relatório, o repertório deve ser limitado.

O Unicode, como sistema de codificação, destina-se a acomodar uma variedade de requisitos para a codificação de texto em linguagem natural (incluindo impressão). O texto em linguagem natural possui várias flexibilidades, incluindo a suposição de que os leitores identificarão erros tipográficos e similares e poderão deduzir o que foi pretendido a partir do contexto.

Identificadores como nomes de domínio geralmente não possuem tal contexto associado. Isso é particularmente verdadeiro para rótulos TLD, nos quais os rótulos na zona de raiz não têm contexto linguístico.²⁰

Por fim, nomes de domínio são utilizados pelos usuários finais. Em circunstâncias normais, precisam ser construídos de maneira a permitir uma transcrição fácil e precisa pelo usuário final de um contexto para outro.

Esses fatores restringem o número e as classes de caracteres Unicode adequados para serem utilizados como parte de um nome de domínio.

4 Recomendações

Recomendação 1: Como os riscos identificados neste comunicado não podem ser adequadamente mitigados sem alterações significativas no Unicode ou no IDNA (ou ambos), o SSAC recomenda que a diretoria da ICANN rejeite qualquer TLD (rótulo de zona de raiz) que inclua *emoji*.

Recomendação 2: Como os riscos identificados neste comunicado não podem ser adequadamente mitigados sem alterações significativas no Unicode ou no IDNA (ou ambos), o SSAC não recomenda o registro de qualquer nome de domínio que inclua *emoji* em qualquer um de seus rótulos. O SSAC também adverte os registradores de nomes de domínio com *emoji* que esses domínios podem não funcionar de forma consistente ou podem não estar acessíveis universalmente como esperado.

5 Agradecimentos, Divulgações de Interesse, Divergências e Retiradas

No interesse da transparência, estas seções fornecem ao leitor informações sobre quatro aspectos do processo do SSAC. A seção Agradecimentos lista os membros do SSAC, especialistas externos e funcionários da ICANN que contribuíram diretamente para este

²⁰ Ver seção 2.1 do SAC060 em <<https://www.icann.org/en/system/files/files/sac-060-en.pdf>>

documento em particular. A seção Divulgações de Interesse aponta para as biografias de todos os membros do SSAC, que divulgam quaisquer interesses que possam representar um conflito - real, aparente ou potencial - com a participação de um membro na preparação deste comunicado. A seção Divergências fornece um local para os membros individuais descreverem qualquer discordância que possam ter ao conteúdo deste documento ou ao processo para prepará-lo. A seção Retiradas identifica membros individuais que se recusaram a discutir o tópico com o qual este comunicado está relacionado. Com exceção dos membros listados nas seções Divergências e Retiradas, este documento tem a aprovação consensual de todos os membros do SSAC.

5.1 Agradecimentos

O comitê deseja agradecer aos seguintes membros do SSAC e aos especialistas externos por seu tempo, contribuições e revisão na produção deste Comunicado.

Membros do SSAC

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

Especialistas convidados

John Klensin
Pete Resnick

Equipe da ICANN

Patrik Jones
Andrew McConachie
Kathy Schnitt
Steve Sheng (editor)

5.2 Divulgações de Interesse

Informações biográficas e Divulgações de Interesse dos membros do SSAC estão disponíveis em: <https://www.icann.org/resources/pages/ssac-biographies-2017-02-16-en>.

5.3 Divergências

Não houve divergências.

5.4 Abstenções

Não houve abstenções