

Uma Introdução à Universal

Acceptance - novos TLDs e internacionalização dos domínios e de e-mail

Sávyo Morais / IX Fórum Regional - RJ
08/11/2019



Universal Acceptance

Exercício de Aquecimento

Cada um dos 3 grupos a seguir contém listas de Top Level Domains (TLDs) válidos (aprovados e delegados pela ICANN), exceto que cada lista possui um que foi inventado ou é inválido. **Qual TLD de cada grupo é inválido?**

Exercício de Aquecimento

Grupo A

嘉里
ANALYTICS
BLOCKBUSTER
DIAMONDS
HOTELES
广东
MOVISTAR
இந்தியா
REALLY
政务

Grupo B

ABC

ATHLETA
இலங்கை
CANCERRESEARCH
CITIC
新加坡
ESURANCE
FAKE
ไทย

Grupo C

GMAIL

JOY
LIKE
hwy
ONYOURSIDE
OOO

SILLY
SUCKS

مصر

قطر

فلسطين

Exercício de Aquecimento

Grupo A

嘉里
ANALYTICS
BLOCKBUSTER
DIAMONDS
HOTELES
广东
MOVISTAR
இந்தியா
REALLY
政务

Grupo B

ABC

ATHLETA
இலங்கை
CANCERRESEARCH
CITIC
新加坡
ESURANCE
FAKE
ไทย

Grupo C

GMAIL

مصر
قطر
JOY
LIKE
hwy
ONYOURSIDE
OOO

فلسطين
SILLY
SUCKS

Testes Práticos

Tente enviar um e-mail para:

savyo@datamail.in

Teste seu próprio endereço de e-mail -

<https://uasg.tech/eai-check/>

Universal Acceptance in a Nutshell

“Pessoas de todo o mundo podem efetivamente usar qualquer nome de domínio e qualquer endereço de e-mail em qualquer aplicação para se uso pessoal ou profissional”

A Internet Evoluiu

2006

Generic

(22 total)

.com	.edu
.org	.net
.aero	.asia
.mobi	.travel
.info	

Country Code

(Mais de 200 total)

.ar	.bo
.br	.co
.cu	.pa
.mx	.uy

2016

Generic

(Mais de 1,200 total)

.com	.edu	.org
.net	.aero	.info
.bank	.bar	.club
.car	.global	.london
.rio	.vote	.sky
.संगठन	.在线	.ابوظبي
.닷넷	.ストア	.дети

Country Code

(Cerca de 300 total)

.ar	.br
.co	.pa
.mx	.uy
.pф	.台灣
.한국	.ଭାରତ

Mas nem tudo vai como o planejado...

UA-ready



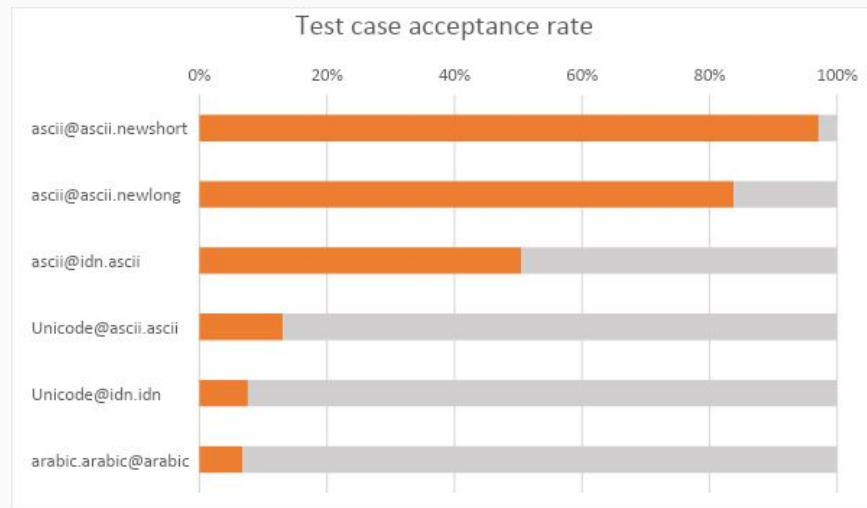
Not ready



Estudo Mundial do UASG

Em 2019, apenas 6% dos 1000 sites mais acessados no mundo* conseguem lidar com nosso conjunto de endereços de e-mail com caracteres não-ASCII e os novos TLDs maiores que os tradicionais dois ou três caracteres ASCII.

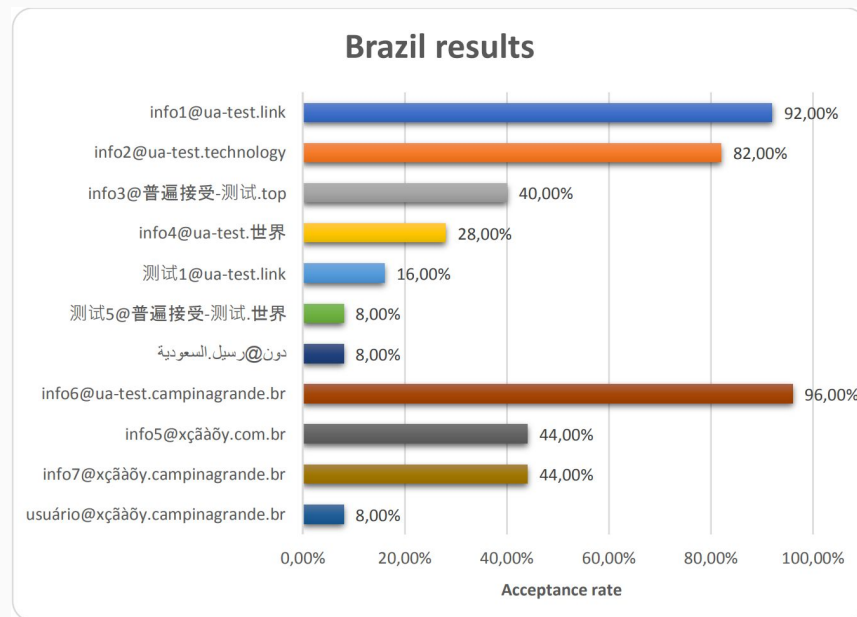
* segundo ranking Alexa



Estudo Brasileiro do UASG

Em 2018, apenas 6% dos 50 sites mais acessados no Brasil* conseguem lidar com nosso conjunto de endereços de e-mail com caracteres não-ASCII e os novos TLDs maiores que os tradicionais dois ou três caracteres ASCII.

* segundo ranking Alexa



Anatomia de um endereço de e-mail

username@example.com

测试5@普遍接受-测试.世界

- * username/ 测试5
 - Nome da caixa de e-mail
 - Pode ser ASCII ou Unicode
- * Nome de domínio de segundo nível e inferiores (example/ 普遍接受-测试)
 - Pode ser ASCII ou Unicode
 - Unicode pode ser representado como Unicode U-label ou ASCII A-label
- * Nome do Top-Level Domain (.com/ .世界)
 - Pode ser ASCII ou Unicode
 - Unicode pode ser representado como Unicode U-label ou ASCII A-label
 - Pode ter de 2 a 63 caracteres
 - Validação no DNS, não por listas estáticas obsoletas

Desafios técnicos à Implementação da Universal Acceptance



SMTPUTF8: o novo SMTP

Por causa da internacionalização dos endereços de e-mail (EAI) criou-se um novo SMTP que funciona em paralelo com a versão legada:

- * Os serviços devem avisar quando suportam SMTPUTF8
- * Há retrocompatibilidade, mas a versão legada não comunica com EAIs
- * Substituir EAIs por Punycode não é válido



Códigos de Validação para E-mail

Expressão regular para cadastro em um grande e-commerce brasileiro ✖

$$/^{\backslash}\mathbf{w}+([\backslash.-]?{\backslash}\mathbf{w}+)^{*}@{\backslash}\mathbf{w}+([\backslash.-]?{\backslash}\mathbf{w}+)^{*}(\backslash.{\backslash}\mathbf{w}\{2,3\})+{\backslash}\$/{}$$

Expressão regular de formulário de contato no site do MEC ✔

$$/^{\wedge}(([\wedge\<\>()\[\]\backslash.,;:\s@\""]+(\backslash.[\wedge\<\>()\[\]\backslash.,;:\s@\""]+)^*)(\backslash\".+\"))@(([\wedge\<\>()\[\]\backslash.,;:\s@\""]+\backslash.)+[\wedge\<\>()\[\]\backslash.,;:\s@\""]\{2,\})\$/{}$$

Atualizando sistemas antigos

- * Em geral não é difícil atualizar um software para a versão mais recente do Unicode, IDNA, SMTP, etc.
 - Pode-se comparar com uma “correção de bug”
- * Os desafios são:
 - Encontrar todas as partes do código que tratam de domínios, URLs, URIs e endereços de e-mail
 - Identificar os casos a serem testados
 - Gerência de strings bi-direcionais
 - “Linkificação”
 - “Não se mexe em time que está ganhando”

HTML5 e EAI

Em sua versão 5, o HTML dispõe da tag `<input type = “email”>` para a inserção de e-mail em formulários. Em 2018 30% dos sites mais acessados do Brasil usavam essa ferramenta para validar e-mails.

A versão 5.2 da linguagem conta com uma definição defasada do que seria um endereço de e-mail válido, excluindo principalmente e-mails com caracteres não ASCII à esquerda do @.

No [draft do HTML5.3](#) isso foi corrigido, mas desde 2017 ainda segue como “First draft”, sem previsão de evoluir para “Candidate recommendation” ou “Recommendation”.

Fique por dentro

- * [Teste](#) seu endereço de e-mail
- * Avalie e corrija seus sistemas
 - * Use o [Blueprint for CIOs](#) do UASG como guia
 - * Encontre [Casos de Uso](#) para testar seu software
- * Inclua as Cláusulas de Preparação da UA em seus contratos
 - * Use as sugestões do UASG disponíveis em [Quick Guide to Tendering](#)
- * Reporte problemas de UA em outras aplicações
 - * UASG [Issue Logging](#)
- * Participe nas discussões do UASG
 - * Entre na [UA Discuss Mailing List](#)
- * Estudos publicados:
 - * 50 sites mais acessados no [Brasil](#) em 2018
 - * 1000 sites mais acessados no [mundo](#) em 2019

savyovm@gmail.com

Dúvidas?

Universal Acceptance
Steering Group

<https://uasg.tech>