



# ¿El Internet es internacional?

## Comprensión de la Aceptación Universal

Mark W. Datysgeld – Governance Primer

2019

# ¿Qué es Aceptación Universal?



Aceptar



Validar



Almacenar



Procesar



Presentar

Internacionalización y competencia:

- ✓ Mayor competencia con **nuevos gTLD**: permitiendo una mayor elección al consumidor en cuanto a los nombres de dominio.
- ✓ Experiencia del usuario en su **idioma local**: dominios con scripts de alfabetos como el **Árabe**, el **Cirílico** y el **Han**.

# Script vs. Lenguaje

**Script:** conjunto de elementos base o símbolos definidos.

A, E, I, O, U ◦ な (Na) に (Ni), ん (Nu), ね (Ne), の (No)

**Idioma:** sistema de comunicación utilizado por una comunidad específica.

**Hola! Qué tal?** ◦ こんにちは、元気ですか?

(Konnichiwa, genkidesuka?)

El **coreano** es un idioma que usa el script **Hangul**.

# Nombres de Dominio Internacionalizados

La diversidad lingüística en el mundo es más fuerte de lo que percibimos en América.

Los IDN se aprobaron por primera vez para su uso en ccTLDs.

Actualmente existen 59 IDN ccTLD aprobados.



# ¿Cómo funciona?

En 2011 se hizo posible la creación de URLs completas en caracteres no ASCII. La tecnología que permite esto es **Punycode**.

Punycode es **Unicode representado en ASCII**.

**Ejemplo (U-Label):** 中国移动.中国 (China Mobile)

**Punycode (A-Label):** xn--fiq02ib9d179b.xn--fiqs8s

**Hazlo tu mismo:**

<https://www.punycoder.com/>

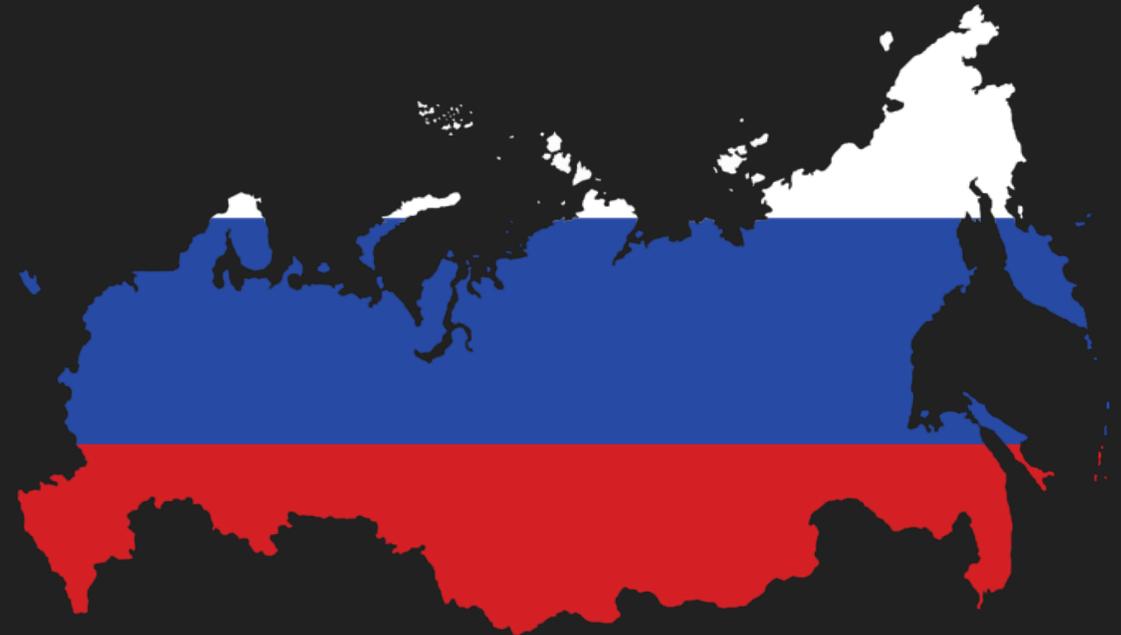
# IDNs: Caracteres cirílicos

**.рф** (Российская Федерация)

Significado: "Federación Rusa"

Más de **1 millón de nombres de dominio**, muchos de ellos legítimos, tales como:

- **МВД.рф**: Ministerio del Interior de Rusia.
- **НОВЫЙДОМ43.рф**: Empresa de ingeniería civil.
- **Геобел.рф**: Proyecto de mapeo de servicios locales.



# IDNs: Caracteres indios <sup>1/2</sup>

**India** tiene:

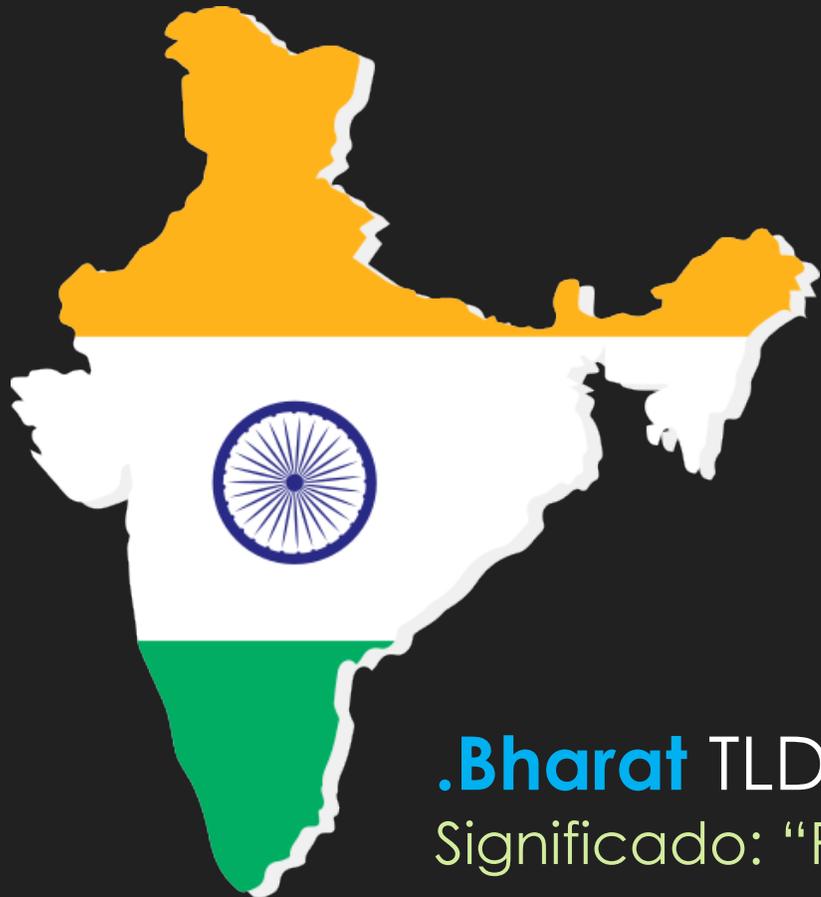
**22** idiomas **oficiales**

**13** alfabetos **oficiales**

**Por ley:**

- 1.** Los teléfonos móviles deben ser compatibles con los idiomas locales.
- 2.** Todos los sitios web oficiales deben estar en inglés y un idioma local.
- 3.** Correos electrónicos en scripts locales son entregados a los ciudadanos por el gobierno.

# IDNs: Caracteres indios <sup>2/2</sup>



**.Bharat** TLD

Significado: "República India"

## IDN

.भारत

.ভারত

.భారత్

.ભારત

.بھارت

.இந்தியா

.भारत

## Idioma

Devanagari

Bengali

Telugu

Gujarati

Urdu

Tamil

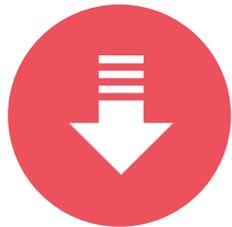
Gurumukhi

# GeoTLDs

abudhabi	brussels	hamburg	moscow	stockholm	zuerich
africa	budapest	helsinki	nagoya	sydney	zulu
africa	capetown	ist	nrw	taipei	МОСКВА
alsace	catalonia	istanbul	nyc	tata	佛山
amsterdam	cologne	joburg	okinawa	tirol	广东
aquitaine	cymru	koeln	osaka	tokyo	广州
bar	doha	kyoto	paris	tui	深圳
barcelona	dubai	london	quebec	vlaanderen	ابوظبي
bayern	durban	madrid	rio	wales	
berlin	fri	melbourne	saarland	wien	
boston	gent	miami	scot	yokohama	

# Campos de e-mail en la Web

En el desarrollo Web, generalmente los nuevos casos de uso del e-mail no se tienen en cuenta al tratarse de **formularios**.



Aceptar



Validar

## Sign up to [ICANN.org](https://www.icann.org)

Create an account on [icann.org](https://www.icann.org) for a more engaging experience. Your account gives you access to new features, including personalized news and policy updates.

Password requirements: 8 characters and must contain at least one small letter, one capital letter, one number and one special character.

By submitting my personal data, I agree that my personal data will be processed in accordance with the [ICANN Privacy Policy](#), and agree to abide by the website [Terms of Service](#).

 I'm not a robot   
reCAPTCHA  
Privacy - Terms

Already have an account? [Log in](#)

# Encuesta global: Premisa <sup>1/3</sup>

**Objetivo:** evaluación del cumplimiento de lo UA en los 1,000 sitios Web más visitados del mundo (según Alexa) mediante el muestreo de enfoques de desarrollo para el campo del e-mail en formularios.

---

ascii@ascii.newshort

test@test.exp

ascii@ascii.newlong

test@test.example

ascii@idn.ascii

test@[普遍接受-测试.org](#)

---

unicode@ascii.ascii

测试1@test.org

unicode@idn.idn

测试5@[普遍接受-测试.世界](#)

arabic.arabic@arabic (RTL)

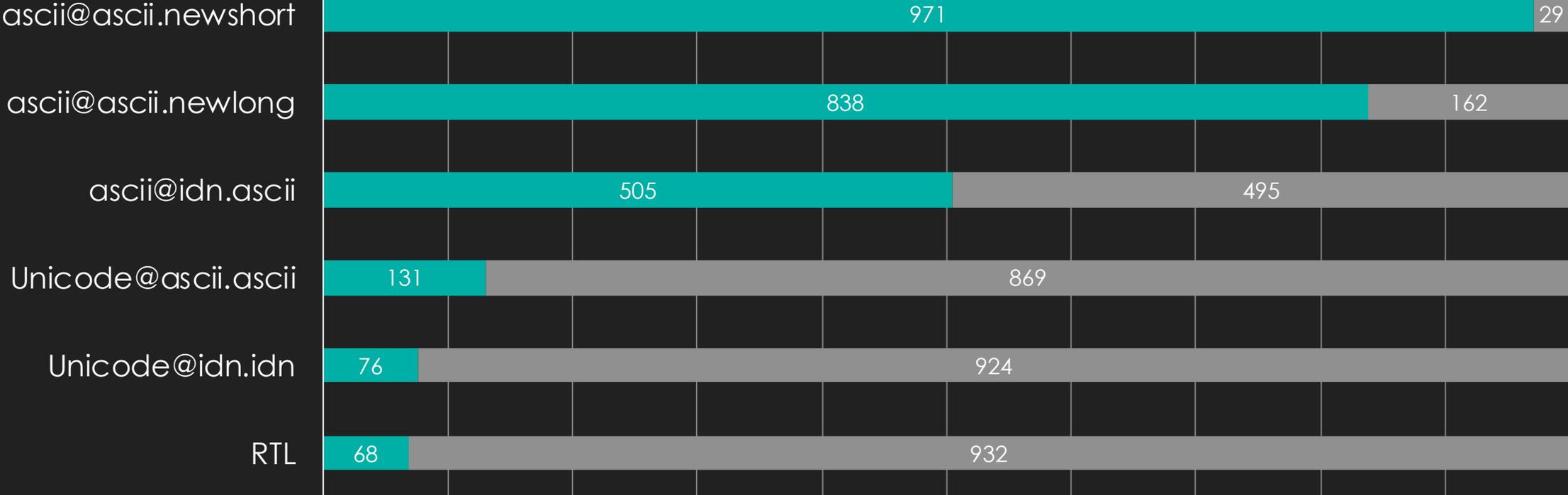
دون@رسيل.السعودية

---

# Encuesta global: Resultados <sup>2/3</sup>

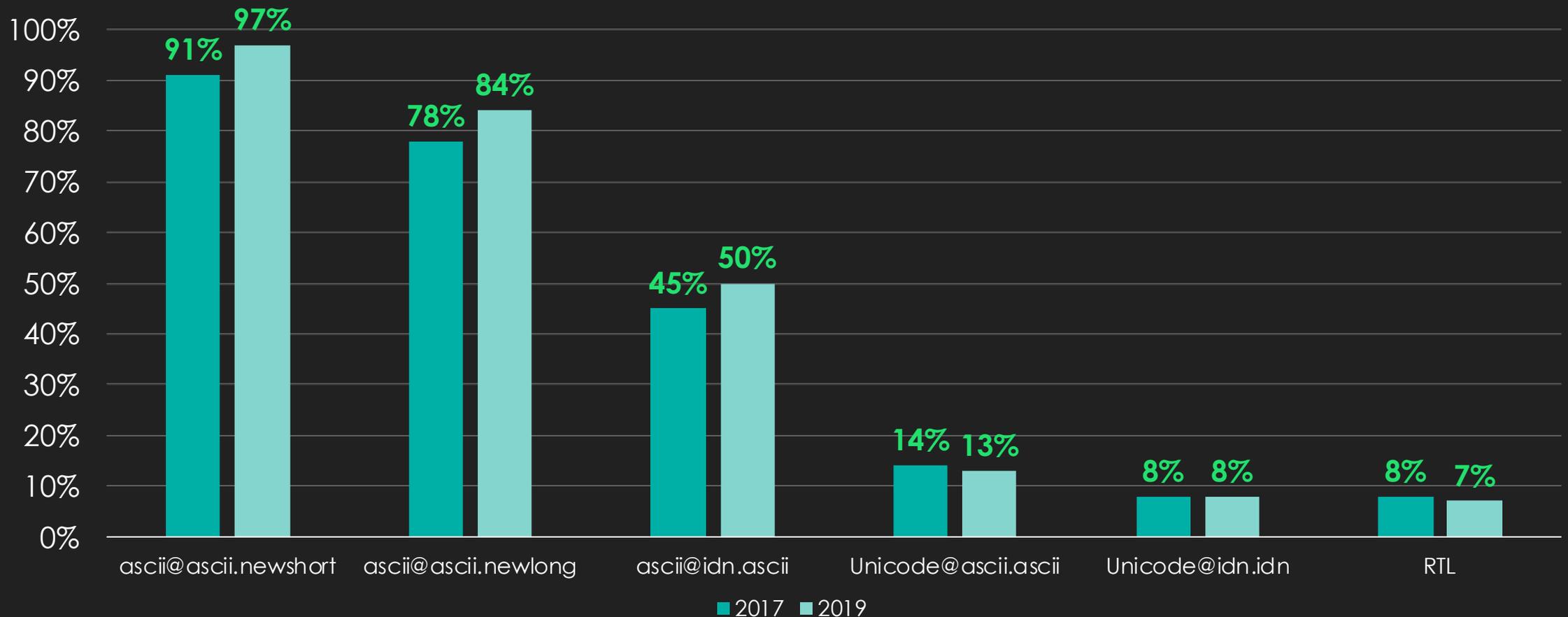
### Tasa de aceptación de caso de prueba

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%



# Encuesta global: Comparativa <sup>3/3</sup>

## Resultados 2017 vs. 2019



# Problemas: Backend

**Email Address Internationalization** (EAI) utiliza caracteres que no son ASCII.

- Exige que SMTP utilice la función SMTPUTF8.
- Los nombres de dominio pueden ser más largos de lo esperado.
- La escritura de derecha a izquierda (RTL) es difícil de entender y mostrar.
- Los “variant characters” se escriben de manera diferente pero significan lo mismo.
- Las pruebas deben hacerse con idiomas desconocidos para los codificadores.

# Problemas: Código obsoleto

## JavaScript Obsoleto

Muchos códigos están hechos para validar solo las letras A a la Z, números del 0 al 9 y “-” (**guion**), que es el estándar ASCII.

Muchos también incluyeron algunos caracteres adicionales tales como “\_” (**guion bajo**).

```
email:function ( email ) {return /^[\w.+\\-]+@ [a-zA-Z0-9-]+ (\\. [a-zA-Z0-9-]+)+$/i.test(email);}
```

Otros van más allá y agregan comprobaciones redundantes para algunos TLD, así como caracteres adicionales:

```
/^[-a-z0-9~!$%^&*_=+}{\'?]+(\\. [-a-z0-9~!$%^&*_=+}{\'?]+)*@([a-z0-9_][-a-z0-9_]* (\\. [-a-z0-9_]+)*\\. (aero|arpa|biz|com|coop|edu|gov|info|int|mil|museum|name|net|org|pro|travel|mobi|[a-z][a-z])|([0-9]{1,3}\\.[0-9]{1,3}\\.[0-9]{1,3}\\.[0-9]{1,3}))(:[0-9]{1,5})?$/i
```

# Problemas: HTML5

Uso: `<input type="email">`

- El 20-30% de los sitios Web utilizan formularios de esta manera.
- El problema: admitir caracteres diferentes de ASCII a la izquierda de la @.
- La propuesta 5.3 de la especificación predice el cumplimiento de UA, pero ha estado en "primera versión" desde 2017.

Objetivo: `<input type="eaimail">`

Caso de prueba	Resultado
ascii@ascii.newshort	Aceptado
ascii@ascii.newlong	Aceptado
ascii@idn.ascii	Aceptado
unicode@ascii.ascii	Rechazado
unicode@idn.idn	Rechazado
RTL	Rechazado

# Problemas: Homógrafos <sup>1/2</sup>

¡Letras similares corresponden a diferentes códigos!

Latín	Griego	Cirilico
H	Η	Н
\x48	\xce\x97	\xd0\x9d

# Problemas: Homógrafos <sup>2/2</sup>

## Caso 2017

Bug en Firefox y Chrome

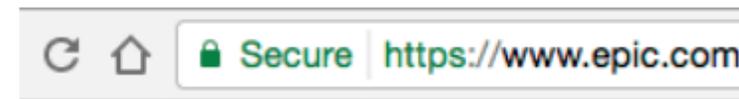
**Sitio Web de ejemplo:** epic.com

**Punycode:** <https://www.xn--e1awd7f.com/>

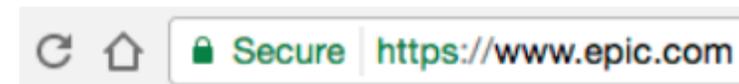
Investigación de WordFence:

<https://www.wordfence.com/blog/2017/04/chrome-firefox-unicode-phishing/>

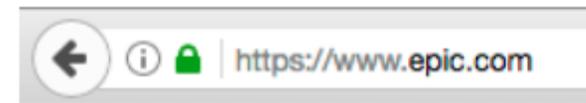
Here is what the real epic.com looks like in Chrome:



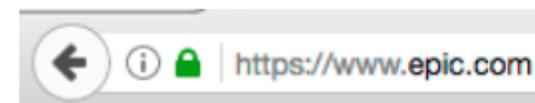
Here is our fake epic.com in Chrome:



And the real epic.com in Firefox:



And here is our fake epic.com in Firefox:



¿Qué hacer?



Universal Acceptance



**¡Gracias!**

**[www.markwd.website](http://www.markwd.website)**