



arsys

La conversión de dominios internacionalizados: Punycode

Qué es punycode, para qué sirve,
ventajas y desventajas de esta
codificación de nombres de dominio



Internet es muy grande y muy diverso y aglutina muchas realidades que reflejar. Hasta el año 2003 era posible crear nombres de dominio solo con caracteres del alfabeto latino, cifras del 0 al 9 y guiones. Esto parecía suficiente, pero la realidad es que en todo el mundo se intentaban utilizar los caracteres propios de sus respectivos alfabetos.

Sin ir más lejos, en español podríamos querer crear un nombre de dominio fiel a nuestra marca si, por ejemplo, nuestra empresa se llamase 'La Piraña', de manera que podríamos pensar en un dominio reconocible por nuestros clientes como «lapiraña.es» (ejemplo ficticio).

Para permitir esto se creó en 2003 el protocolo de nombres de dominio internacionalizados en aplicaciones (IDNA) que establece una traducción estandarizada unívoca del código Unicode al código ASCII.

Gracias a este protocolo es posible crear y utilizar dominios con cualquier carácter Unicode, algo que va a beneficiar directamente al marketing de las empresas que, por ejemplo, se planteen un ecommerce. Tan solo queda por entender cómo casa este dominio Unicode con Internet, que entiende básicamente código ASCII.

Punycode, el traductor de dominios

En esencia, **Punycode** es un método de codificación estándar que hace ese trabajo que mencionamos anteriormente: permite expresar caracteres Unicode usando el conjunto (más limitado en número) de caracteres ASCII. Como dijimos al principio, los dominios aceptan caracteres ASCII como las letras minúsculas (de la «a» a la «z»), los dígitos del 0 al 9 y el guion («-»).

Hablamos de los caracteres básicos, así que no podemos incluir la «ñ» en ese conjunto. Se utiliza **Punycode** como el elemento traductor capaz de procesar los nombres de dominio internacionalizados y expresarlos directamente con ese pequeño conjunto de caracteres aceptados.

Los IDN son los nombres de dominio internacionalizados que aceptan, además de la ñ, caracteres como la diéresis, la cedilla o la tilde, además de otras letras no originales del alfabeto latino. Podríamos formar un dominio con «cigüeña», por ejemplo.

Punycode es una aplicación de un algoritmo de codificación más general que se llama Bootstring, que permite representar cadenas de caracteres con un conjunto más reducido o limitado de elementos. Para hacerlo, se basa en seis principios:

- **Integridad:** cada cadena de salida se representa por una cadena simplificada.
- **Univocidad:** cada cadena de salida se representa exactamente por una sola cadena simplificada.
- **Reversibilidad:** el proceso inverso de conversión se realiza sin pérdida de información.

- **Eficiencia:** la cadena de caracteres codificada supera mínimamente en longitud a la cadena de salida.

- **Sencillez.**

- **Legibilidad:** tan solo se codifican los caracteres no representables en el conjunto de caracteres de destino.

Para convertir un dominio con un carácter especial como la «ñ» se utiliza el conocido como prefijo ACE (xn--), que da lugar a unos nombres de dominio que puede que te suenen gracias a ese comienzo tan poco común.

En nuestro ejemplo anterior, el dominio lapiraña.es se leería, aplicando ACE, como xn--lapiraa-9za.es.

Obviamente, visto así es poco atractivo. Sin embargo, dentro de las fronteras de nuestro país y en cualquier lugar de habla hispana, utilizar un dominio como lapiraña.es puede ser ventajoso en cuestiones de marketing.

Existen multitud de conversores gratuitos de dominios de este tipo para conocer la versión representada por ASCII, aunque en el RFC 3492 se describe en profundidad el algoritmo para que se pueda realizar una implementación del procedimiento de codificación utilizando el lenguaje C.



Punycode y el cibercrimen

Hay un cierto riesgo asociado a Punycode. Existe un tipo de ciberataque denominado homográfico que se utiliza para el phishing, y que se aprovecha de Punycode para terminar de engañar al usuario.

Al poder utilizar caracteres Unicode, un atacante puede elegir una letra que se parezca mucho a la que pretende sustituir. Ni siquiera tenemos que pensar en Unicode, simplemente, pensemos en la diferencia real entre estos dos caracteres: «l», «l». ¿Cuál es una e minúscula y cuál es una i mayúscula?

Al extender las posibilidades mediante Punycode un atacante puede sustituir la letra a en un dominio por el carácter cirílico «a» (U+0430). A simple vista, nos va a ser imposible detectar la diferencia. Este error lo detectó Xudong Zheng y lo explicó en su blog.

Los ataques homográficos no son exclusivos de esta codificación, ya que hemos visto que podemos ponerlos en práctica con cualquier otro carácter. Sin embargo, utilizando dominios Punycode aumentamos las probabilidades de picar ante un ataque de phishing.

En Arsys te podemos asesorar y guiar en el proceso de creación de una tienda online gracias a nuestros servicios de ecommerce: Tienda Online Básica, Avanzada y Premium, y también con un Hosting WP + WooCommerce. Te podrás beneficiar de un montón de ventajas, desde la facilidad de uso, pasando por la optimización SEO, todos los estándares de seguridad y soporte técnico las 24 horas.

arsys

www.arsys.es

 www.facebook.com/arsys.es

 twitter.com/arsys

 www.linkedin.com/company/arsys-internet/

