

Aspectos fundamentales: Unicode, IDNs, EAI, UA

Plan

- Aspectos fundamentales clave (solo los conceptos necesarios)
 - Unicode, UTF-8, Normalization
 - Nombres de dominio, etiquetas, dominios de nivel superior (TLD), zonas
 - Nombres de dominio internacionalizados (IDNs): Punycode, U-label, A-label.
 - Lista de Sufijos Público (Public Suffix List, PSL)
 - Internacionalización de direcciones de correo electrónico (Email Address Internationalization, EAI)
 - Aceptación universal (Universal Acceptance, UA)

Nota

- Esta presentación es un resumen sobre múltiples temas. Tiene por objective dar conceptos fundamentales.
- Tengan en cuenta que cada tema se puede cubrir en un tutorial de día completo, por lo que este resumen se mantiene simples y no puede considerarse completos.

Unicode

- Codificación de glifos (glyphs) en caracteres (code points) .
- En las especificaciones, los caracteres se muestran en notación hexadecimal, por ejemplo U+0061 (letra minúscula A).
- Los caracteres normalmente se transportan utilizando el formato UTF-8 (Unicode Transformation Format, 8 bits).
 - Número variable de bytes para un solo carácter.
 - ASCII se usa tal cual.
 - Estándar para transportar caracteres Unicode en la web, protocolos, etc.

Unicode

- Hay varias formas de usar un glifo:
 - “è” = U+00E8
 - “e`” = “è” = U+0065 U+0300
 - La normalización es un proceso para garantizar que, independientemente de lo que el usuario tipee (input), la representación final sea la misma.
 - Para las dos entradas anteriores, la forma normalización C (NFC) generará U+00E8 para ambos

Nombres de Dominio Internacionalizados (IDNs)

- Los nombres de dominio internacionalizados (IDN) permiten el uso de caracteres no ASCII (e.g. fuera del rango A-Z) para cualquier etiqueta de un nombre de dominio.
 - No todas las etiquetas de un nombre de dominio necesitan estar internacionalizadas.
- Ejemplo: exâmples.ca
- El usuario usa la versión IDN, pero el IDN se convierte en ASCII para la resolución de DNS.
 - exâmples => exmples-xta => xn--exmples-xta
 - El prefijo 'xn--' se agrega para identificar un IDN.

Nombres de Dominio Internacionalizados (IDNs)

- Proceso de ejemplo de uso de un IDN:
 - El usuario ingresa en un navegador: `http://exâmples.ca`
 - El navegador realiza la normalización en la entrada del usuario.
 - El navegador convierte `exâmples.ca` en una representación compatible con ASCII por medio del algoritmo Punycode [RFC3492] y agrega "xn--" delante de él.
 - *exâmples.ca* se convierte en: *xn--exmple-xta.ca*
 - El navegador llama al DNS para obtener la dirección IP de *xn--exmple-xta.ca*
 - El navegador se conecta al servidor HTTP en la dirección IP recibida.

Nombres de Dominio Internacionalizados (IDNs)

- El protocolo para manejar IDN se denomina IDN para Aplicaciones (IDNA).
 - Dos versiones: IDNA2003 e IDNA2008. Este último (IDNA2008) es el que se utiliza actualmente.
- *U-label* es la representación nativa Unicode de una etiqueta IDN: *exâmple*
- *A-label* es la representación de una etiqueta IDN en formato ASCII: *xn--exmple-xta*
- La conversión entre A-Label y U-label es reversible.

Dos estándares de IDN

- IDNA2003 (RFC3490)
 - Se definió con Unicode 3.2 (marzo de 2002).
 - Los caracteres más nuevos definidos después de Unicode 3.2 se aceptan tal cual.
- IDNA2008
 - Basado en propiedades de puntos de código Unicode.
 - Los nuevos caracteres son manejados automáticamente por sus propiedades.
 - Por lo tanto, se acepta un menor número de caracteres en IDNA2008 que en IDNA2003.

- Sintaxis del correo electrónico:
nombre_buzón@nombrededominio
- El nombre de dominio se puede internacionalizar como un IDN
- Cuando la parte local o nombre de buzón usa Unicode (UTF8) es EAI.
- Ejemplos:
 - k evin@example.org
 - すし @ 快手. 游戏

- Efecto secundario:
 - Los encabezados de correo deben actualizarse para admitir EAI.
 - El software de correo utiliza los encabezados de correo para obtener más información sobre cómo entregar el correo electrónico.
- Dado que no todos los servidores de correo electrónico admiten EAI, se utiliza un protocolo de negociación para enviar EAI solo cuando el servidor de destino lo admite. De lo contrario, retrocede y devuelve un mensaje que no se puede entregar al remitente.
- La opción SMTPUTF8 se utiliza dentro del protocolo de transferencia de correo (SMTP: Protocolo simple de transporte de correo) para señalar el soporte de EAI por un servidor de correo electrónico.

Aceptación Universal (UA)

- UA propone que todos los nombres de dominio y correo electrónicos se procesen satisfactoriamente para dar una buena experiencia al usuario final.
 - Los nombres de dominio o correos electrónicos pueden usar cualquier sistema de escritura: Español, Chino, Árabe, etc.
 - Los nombres de dominio de alto nivel (Top Level Domain) pueden tener cualquier longitud (hasta 63 octetos)

Aceptación Universal (UA), continuación

- UA también se trata de:
 - TLD de cadena nuevos y más largos (nuevos dominios genéricos de nivel superior - nuevos gTLD)
 - En los primeros días, los TLD tenían dos o tres caracteres (.ca, .com). Recientemente, los TLD comenzaron a tener cadenas más largas (.info, .google)
 - Una etiqueta de TLD puede tener hasta 63 octetos.
 - Algunas aplicaciones aún están verificando que el TLD ingresado por un usuario tenga un máximo de 3 caracteres.
 - Otros están usando expresiones regulares que toman un máximo de 6-7 caracteres para el TLD

Aceptación Universal (UA)

- UA también se trata de:
 - TLD agregados / eliminados (de la raíz del DNS):
 - Los TLD van y vienen a diario.
 - Algunas aplicaciones están verificando la corrección de un TLD basándose en una lista estática que no es la más reciente, por lo que hacen suposiciones erróneas sobre la existencia de un TLD.

Lista de sufijos públicos (PSL)

- La Public Suffix List (PSL) es un intento de ayudar a los desarrolladores web a saber si los dominios están controlados por la misma organización. Esto se puede considerar como una pregunta: ¿este nombre de dominio permite que otros se registren con él?
- <https://publicsuffix.org/>
- Algunas bibliotecas, marcos y aplicaciones utilizan PSL como una lista estática de TLD.
- Si se usa en aplicaciones:
 - El desarrollador debe mantener la copia actualizada en su aplicación.
 - Es posible que un TLD no esté incluido en el PSL.
- Para saber si un nombre es realmente un TLD, utilice otros medios.