



Guía rápida sobre Aceptación Universal



ACEPTAR



VALIDAR



ALMACENAR



PROCESAR



VISUALIZAR

El software y los servicios en línea son compatibles con la Aceptación Universal cuando ofrecen las capacidades enumeradas previamente para todos los dominios y todas las direcciones de correo electrónico. electrónicos.

¹El sistema ASCII es la codificación de caracteres utilizada históricamente en Internet, definida en la norma de Internet RFC 20 (<https://tools.ietf.org/html/rfc20>). La definición del sistema Unicode le compete al Consorcio Unicode (<http://unicode.org>).

¿Qué significa “Aceptación Universal”?

Algunos programas informáticos no reconocen o procesan correctamente todos los nombres de dominio y todas las direcciones de correo electrónico. Los nombres de dominio pueden incluir cadenas de caracteres en el nivel superior que sean más largas que las más tradicionales y conocidas, y los nombres de dominio y las direcciones de correo electrónico ahora pueden utilizar caracteres extraídos de un repertorio de Unicode mucho más amplio que el código ASCII tradicional¹. La **Aceptación Universal** (UA) es el estado en el cual todos los nombres de dominio y las direcciones de correo electrónico válidas se **aceptan, validan, almacenan, procesan** y **visualizan** de manera correcta y uniforme.

El Grupo Directivo sobre Aceptación Universal (UASG) es una iniciativa liderada por la comunidad cuyo objetivo es generar conciencia, identificar y resolver los problemas asociados con la aceptación universal de todos los nombres de dominio y las direcciones de correo electrónico. El propósito de esta iniciativa es ayudar a garantizar una experiencia uniforme y positiva para los usuarios de Internet a nivel global. Cuenta con el apoyo de la ICANN (Corporación para la Asignación de Nombres y Números en Internet) e incluye a participantes de más de 200 organizaciones de todo el mundo como Afiliadas, Apple, CNNIC, GoDaddy, Google, Microsoft y Verisign, entre otras. Para más información sobre el UASG y sus avances recientes, los invitamos a visitar www.uasg.tech.

Esta **Guía Rápida** describe las recomendaciones del UASG para lograr la Aceptación Universal en los cinco contextos (aceptar, validar, almacenar, procesar y visualizar) en los que los sistemas encuentran nombres de dominio y direcciones de correo electrónico. Está dirigida a ejecutivos y gerentes responsables de las actividades de Tecnología de la Información o Ingeniería de Productos de Software. Presenta las recomendaciones del UASG a nivel general sin algunos de los detalles que serían importantes para un arquitecto o ingeniero en software. Para conocer esos detalles, les recomendamos consultar el documento UASG 007, "Introducción a la Aceptación Universal".

ACEPTAR



La **aceptación** es el proceso por el cual una dirección de correo electrónico o un nombre de dominio es recibido desde una interfaz del usuario, un archivo o una API (Interfaz de Programación de Aplicaciones) para su utilización en una aplicación de software o un servicio en línea.

Recomendaciones del UASG

- Los campos de entrada deberían ser lo suficientemente grandes como para aceptar cualquier entrada válida. En función de cómo esté codificado, un nombre de dominio puede requerir hasta 670 bytes. Una dirección de correo electrónico puede tener una parte local (la parte antes del signo @) de hasta 64 bytes además de un nombre de dominio, con una longitud total de hasta 735 bytes.
- Las aplicaciones y servicios deberían aceptar nombres de dominio y direcciones de correo electrónico codificados con UTF-8² y deberían reconocer que la cantidad de bytes ocupados por la codificación con UTF-8 puede ser mayor que la cantidad de caracteres visualizados.
- Se puede ingresar y visualizar un IDN en su código de escritura original o en una versión ASCII diseñada para que sea compatible con versiones previas; por ejemplo: 测试 y xn--0zwm56d. La codificación Unicode del código de escritura original se denomina Etiqueta U; la codificación equivalente compatible con ASCII se denomina Etiqueta A.³ El software debería aceptar tanto las etiquetas A como las etiquetas U, en lugar de convertir las etiquetas A en etiquetas U para su visualización y para cualquier procesamiento que no requiera etiquetas A.
- En casi todos los casos, un nombre de dominio o una dirección de correo electrónico que se ingresa debería convertirse a la Forma de Normalización de Unicode C (NFC)⁴ antes de su procesamiento posterior. Debido a que la NFC no es completamente a prueba de pérdidas, en circunstancias excepcionales puede ser necesario diferir la normalización hasta que el procesamiento posterior haya establecido el contexto o contextos específicos en los cuales debería aplicarse.

²El UTF-8 codifica cada punto de código Unicode como una secuencia de uno a cuatro bytes. Se lo define en el documento RFC 3629.

³La conversión entre etiquetas U y etiquetas A se realiza mediante el algoritmo "Punycode" que se define en los documentos RFC 3492 y RFC 5891.

⁴Ver el Anexo de la Norma Técnica 15 de Unicode: "Formas de normalización de Unicode" (<https://www.unicode.org/reports/tr15/tr15-47.html>).

VALIDAR



La **validación** es el proceso para comprobar la sintaxis correcta de una dirección de correo electrónico o nombre de dominio y, cuando corresponda, la existencia de un nombre en el DNS. Es posible que sea necesario actualizar las técnicas de validación para que funcionen con direcciones de correo electrónico y nombres de dominio actuales.

⁵En los documentos RFC 5890, 5891, 5892, 5893 y 5894 se puede consultar la definición de la IDNA2008.

⁶Ver la lista de dominios de alto nivel en <https://www.icann.org/resources/pages/tlds-2012-02-25-en>.

Recomendaciones del UASG

- Los datos de entrada deberían validarse de manera apropiada según el uso al cual están destinados. Todos los nombres de dominio deberían validarse según la norma de Nombres de Dominio Internacionalizados en Aplicaciones vigente (IDNA2008).⁵ Esto asegura la validez sintáctica del nombre.
- Si se espera que una cadena de caracteres ingresada sea una entrada existente en el DNS, es necesario validarla mediante una búsqueda en el DNS.
- Si se espera que una cadena de caracteres ingresada sea un nombre de dominio válido que podría no estar (aún) en el DNS, todavía puede ser posible validarla en forma parcial. Por ejemplo, el nombre de dominio de alto nivel (TLD) se puede verificar con la lista autoritativa de nombres de TLD válidos que mantiene la Autoridad de Números Asignados en Internet (IANA).⁶
- Para validar una dirección de correo electrónico, hay que validar la parte del dominio como se describe anteriormente. Debido a que la parte local de una dirección de correo electrónico está definida únicamente por el sistema de correo electrónico que recibe el correo, generalmente no es posible validarla. Si se solicita al usuario que introduzca la dirección de correo electrónico dos veces, se pueden detectar errores de escritura.
- En la mayoría de los casos, todos los componentes de un nombre de dominio o dirección de correo electrónico (excepto el nombre del TLD si no es un IDN) deberían estar en un único código de escritura (por ejemplo, árabe o han) o en códigos de escritura estrechamente relacionados (por ejemplo, kanji japonés, katakana, hiragana y romaji). Consultar la Norma Técnica 39 de Unicode, "Mecanismos de seguridad de Unicode" (https://unicode.org/reports/tr39/#Restriction_Level_Detection), para verificar que los códigos de escritura en una secuencia de Unicode sigan las prácticas recomendadas.

ALMACENAR



El **almacenamiento** se refiere a almacenar de forma temporaria o a largo plazo nombres de dominio y direcciones de correo electrónico, que deberían almacenarse en formatos bien definidos independientemente de la duración prevista del almacenamiento.

Recomendaciones del UASG

- En casi todos los casos, los nombres de dominio y las direcciones de correo electrónico deberían normalizarse según la Forma de Normalización de Unicode C (NFC) antes de su almacenamiento. Debido a que la NFC no es completamente a prueba de pérdidas, en circunstancias excepcionales puede ser necesario diferir la normalización hasta que el procesamiento posterior haya establecido el contexto o contextos específicos en los cuales debería aplicarse.
- En la mayoría de las aplicaciones, los nombres de dominio y las direcciones de correo electrónico deberían almacenarse en archivos y bases de datos codificados como UTF-8, la codificación Unicode más común y con mejor compatibilidad. En algunos casos, cuando el software tiene que interoperar con bases de datos legadas, puede resultar más fácil utilizar la misma codificación de la base de datos.
- Dentro del código de la aplicación, la representación más apropiada de Unicode depende del entorno de programación. Muchos lenguajes de programación comunes, incluidos los lenguajes Python y Perl, tienen soporte incorporado para Unicode y conversión automática hacia o desde UTF-8 en entrada y salida.
- Las aplicaciones deberían elegir una representación interna uniforme, ya sea en etiquetas U o etiquetas A, para los IDN. Dado que cada etiqueta U puede transformarse en una etiqueta A única y viceversa, cualquiera de las dos formas es aceptable.

PROCESAR



El **procesamiento** tiene lugar siempre que una aplicación o un servicio utiliza una dirección de correo electrónico o un nombre de dominio para realizar una actividad (por ejemplo, realizar una búsqueda o clasificar una lista) o se cambia a un formato alternativo (por ejemplo, de una codificación legada a UTF-8). Puede realizarse una validación adicional durante el procesamiento.

Recomendaciones del UASG

- A medida que Unicode evoluciona, resulta necesario actualizar el software cuando sea posible para utilizar la versión más reciente de la norma y todos los gráficos y fuentes disponibles. Se debe tener en cuenta que los dispositivos de usuario, las bibliotecas de software y los estándares web pueden no ser compatibles con la versión más reciente y, por lo tanto, pueden mostrar los nuevos caracteres asignados incorrectamente, como un cuadro genérico (□), o no mostrarlos en absoluto.
- Cuando las API que admitan entrada o salida de UTF-8 se encuentren disponibles, se las debe utilizar en lugar de las API que no las admitan. Se deben usar bibliotecas estándar bien depuradas, como la GNU libidn2 (<https://www.gnu.org/software/libidn/#libidn2>), para procesar y validar los IDN; no “implementar bibliotecas propias”.
- Los códigos de escritura que se escriben de derecha a izquierda requieren consideraciones especiales cuando se utilizan en nombres de dominio y direcciones de correo electrónico. Algunas de estas consideraciones se abordan en el documento sobre IDNA⁷ (para nombres de dominio) y en un Anexo del Estándar Unicode⁸ (para direcciones de correo electrónico).
- Al crear registros u otras estructuras de datos que incluyan información sobre códigos de escritura o idiomas, se recomienda incluir la mayor cantidad posible, preferentemente todos los que admita el Estándar Unicode.⁹ Hay que tener presente que algunos idiomas se pueden escribir en distintos códigos de escritura, y que algunos códigos de escritura se pueden usar para escribir en una gran cantidad de idiomas distintos.

⁷Ver el documento RFC 5893, “Códigos de escritura de derecha a izquierda para Nombres de Dominio Internacionalizados en Aplicaciones (IDNA)” (<https://tools.ietf.org/html/rfc5893>).

⁸Ver UAX#9, “Algoritmo bidireccional de Unicode” (<http://unicode.org/reports/tr9>).

⁹Ver el documento de Unicode “Códigos de escritura admitidos” (<http://unicode.org/standard/supported.html>).

VISUALIZAR



La **visualización** tiene lugar siempre que una interfaz del usuario represente visualmente una dirección de correo electrónico o un nombre de dominio. La visualización de nombres de dominio y direcciones de correo electrónico suele ser sencilla cuando los códigos de escritura utilizados y los mecanismos de representación necesarios son compatibles con el sistema operativo subyacente y las cadenas de caracteres se almacenan en una codificación definida por el Estándar Unicode. De lo contrario, es posible que se requieran transformaciones específicas para cada aplicación.

Recomendaciones del UASG

- Hay que tener en cuenta que aunque el software y los dispositivos modernos pueden mostrar casi todos los puntos de código Unicode, los sistemas más antiguos pueden tener una compatibilidad limitada y probablemente requieran que las aplicaciones gestionen algunas de sus fuentes antiguas. Además, cuando Unicode agrega nuevos puntos de código, los dispositivos y las aplicaciones no los mostrarán hasta que sus bibliotecas de fuentes se hayan actualizado.
- Los IDN se deben mostrar con sus caracteres originales a menos que exista un requisito específico para mostrarlos como etiquetas A.
- Los nombres de dominio y las direcciones de correo electrónico se pueden visualizar en texto de izquierda a derecha (LTR), como en inglés o ruso, o de derecha a izquierda (RTL), como en árabe o hebreo. Debido a que Unicode asigna atributos de direccionalidad a puntos de código individuales (no a secuencias de puntos de código), no todos los textos LTR y RTL integrados ("bidireccionales") tienen sentido para los usuarios. Es necesario aplicar los criterios¹⁰ de niveles de restricción de Unicode para señalar cadenas de caracteres potencialmente confusas.
- Los usuarios de Internet leen y se comunican en una gran cantidad de idiomas diferentes. En algunos casos puede ser necesario diseñar aplicaciones por separado para diferentes idiomas o grupos de idiomas.

¹⁰Ver la Norma Técnica de 39 de Unicode, "Mecanismos de seguridad de Unicode"

(https://www.unicode.org/reports/tr39/#Restriction_Level_Detection), para consultar sus niveles de restricción moderadamente restrictivo y altamente restrictivo, y verificar que los códigos de escritura en una secuencia Unicode sigan las prácticas recomendadas.

Cómo prepararse para la Aceptación Universal

Revisión de códigos fuente y pruebas de sistemas

Es necesario revisar y evaluar el software y los sistemas desarrollados o actualizados para ser compatibles con la Aceptación Universal a fin de corroborar su buen funcionamiento, como también detectar y corregir errores. Como parte de sus iniciativas de concientización, el UASG se está comunicando con los desarrolladores de aplicaciones y los proveedores de servicios en línea a fin de incentivarlos a realizar revisiones y pruebas de código fuente de Aceptación Universal, y compartir una lista de criterios que puedan usarse para desarrollar casos de prueba estandarizados.

Pruebas

El UASG también está elaborando una lista de sitios web, aplicaciones, direcciones de correo electrónico y nombres de dominio aptos para la realización de pruebas. En algunos casos, las pruebas pueden automatizarse y ejecutarse sin intervención manual. Un ejemplo real es la reciente investigación sobre gTLD que realizaron los laboratorios del APNIC en representación de la ICANN: <https://tinyurl.com/new-gtld-ua>. El UASG está investigando métodos de pruebas automatizadas para la Aceptación Universal y compartirá sus conclusiones cuando estén disponibles.

Más información

En estos documentos se puede consultar más información sobre Aceptación Universal, Unicode y Nombres de Dominio Internacionalizados.

- ▶ UASG 007, “Introducción a la Aceptación Universal” (<https://uasg.tech/documents>).
- ▶ RFC 5894, “Nombres de Dominio Internacionalizados en Aplicaciones (IDNA): antecedentes, explicación y fundamento” (<https://www.rfc-editor.org/info/rfc5894>).
- ▶ “Tipografía internacional en la Web”, resumen gráfico que muestra los problemas y cuestiones en relación al uso de diferentes idiomas en la Web (<https://w3c.github.io/typography/gap-analysis/language-matrix.html>).

Nota sobre terminología

Una de las dificultades para lograr la Aceptación Universal es que muchos términos y conceptos que son familiares para las personas acostumbradas a códigos de escritura simples con una pequeña cantidad de caracteres "alfabéticos" distintos, como el alfabeto latino, pueden crear una gran confusión cuando se aplican a sistemas de escritura con principios diferentes. Para poder incorporar una amplia variedad de sistemas de escritura al campo de los nombres de dominio internacionalizados (IDN), fue necesario crear nuevos términos, como también usar términos familiares (como "carácter") de maneras especiales y muy específicas. Esta Guía Rápida intenta evitar dichos términos, o definirlos cuando se los utiliza. Sin embargo, es probable que sea necesario un conocimiento más profundo de la terminología para analizar otros materiales, incluidos algunos de los documentos mencionados en esta guía.

