

# Evaluación global: aceptación de correos electrónicos en sitios web en 2019

Por el Grupo Directivo sobre Aceptación Universal (UASG),  
la Asociación Brasileña de Empresas de Software (ABES),  
y Governance Primer

9 de agosto de 2019



## ÍNDICE

<b>Introducción</b>	<b>3</b>
<b>Resumen Ejecutivo</b>	<b>4</b>
<b>Integrantes del equipo de investigación</b>	<b>5</b>
<b>Metodología</b>	<b>6</b>
Procedimientos	6
Esquema de metadatos	9
<b>Resultados</b>	<b>11</b>
<b>Evaluación y correlaciones</b>	<b>13</b>
Errores de validación de la muestra	13
Rechazo de todas las direcciones de correo electrónico	13
Pruebas de correlación regional	14
<b>HTML5: el obstáculo para la Aceptación Universal</b>	<b>15</b>
<b>¿Pueden automatizarse estos informes?</b>	<b>17</b>
<b>Conclusión</b>	<b>18</b>



## Introducción

El objetivo de la Aceptación Universal (UA) es garantizar que todos los nombres de dominio y las direcciones de correo electrónico puedan ser utilizados en todos los dispositivos, sistemas y aplicaciones habilitados para funcionar en Internet. Esto contempla tanto los nuevos dominios genéricos de alto nivel (gTLD) como los dominios que no utilizan el código de escritura latino. Si bien algunas personas suponen que estos funcionan de la misma manera que los dominios legados, no es ese el caso. Además, los problemas de compatibilidad siguen siendo más comunes de lo que deberían ser.

Nuestro objetivo es que este correo electrónico:  
测试1@server.technology

Y este correo electrónico:  
دون@رسيل.السعودية

Tengan la misma tasa de aceptación que este:  
**user@test.org**

Este estudio fue encargado por el Grupo Directivo sobre Aceptación Universal (UASG) como seguimiento de una prueba<sup>1</sup> similar realizada en 2017, como parte de una iniciativa más amplia para que la comunidad comprenda mejor los problemas y las cuestiones clave relacionados con la compatibilidad generalizada de todos los nombres de dominio disponibles actualmente.

El objetivo fue evaluar el cumplimiento de la Aceptación Universal en los 1000 sitios web más importantes del mundo (según Alexa<sup>2</sup>) mediante un muestreo de las prácticas y los diferentes enfoques de desarrollo del campo del correo electrónico en formularios web y su puesta a prueba en la práctica. Estos formularios de correo electrónico son adecuados para probar diversos aspectos de la Aceptación Universal, dado que se pueden comprobar diferentes puntos de fallas, incluidos los estándares e implementación de HTML, así como el uso de JavaScript y otros lenguajes orientados al entorno web.

Es evidente que algunos desarrolladores no tienen en cuenta los nuevos casos de uso. Como consecuencia, las personas que desean utilizar direcciones innovadoras o ser capaces de escribir en su propio idioma, necesitan tener a mano un correo electrónico de emergencia para cuando su dirección preferida inevitablemente no pueda ser procesada. Es necesario comprender dónde se encuentran estos problemas para poder resolverlos en forma definitiva.

El resto de este informe proporcionará información sobre la metodología del estudio y describirá los resultados, problemas y posibles soluciones.

---

<sup>1</sup> <https://uasg.tech/wp-content/uploads/2017/09/UASG-Report-UASG017.pdf>

<sup>2</sup> <https://www.alexa.com/topsites>



## Resumen Ejecutivo

Después de filtrar los dominios falsos, 527 de los 1000 mejores sitios web recopilados de la lista de Alexa resultaron ser comprobables. El equipo continuó con la búsqueda de sitios web comprobables hasta el puesto 1922 de la lista para poder obtener mil dominios comprobables. Luego se realizaron evaluaciones manuales para verificar la aceptación de los siguientes formatos de direcciones en sus formularios, ordenados por grado de complejidad:

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

Luego, los resultados se asignaron a una base de datos de MongoDB para facilitar la referencia en un esquema de metadatos creado durante el desarrollo de la metodología; los resultados presentados en el informe derivan de ese trabajo. Si se tiene en cuenta la diferencia de aceptación entre las pruebas de 2019 y 2017, podemos observar un patrón alentador de mayor aceptación para una parte de la misión de la Aceptación Universal, como se puede observar a continuación:

Caso de prueba	2017	2019
ascii@ascii.newshort	91 %	97%
ascii@ascii.newlong	78%	84%
ascii@idn.ascii	45%	50%
unicode@ascii.ascii	14%	13%
unicode@idn.idn	8%	8%
arabic.arabic@arabic (RTL)	8%	7%

En parte, la variación puede explicarse por el uso de diferentes conjuntos de datos. Además, en cada prueba se utilizó la lista de los principales sitios web tal como existía durante la fase de recopilación de datos de cada proceso de evaluación. En este caso, puede ser que la aceptación de algunos casos de prueba no esté disminuyendo, sino que se haya quedado estancada. Sin embargo, el aumento de la aceptación de nuevos dominios y Nombres de Dominio Internacionalizados (IDN) en el segundo nivel muestra un progreso inequívoco y es un asunto de gran interés.

El problema más significativo que encontramos fue `<input type="email">`, que es el campo de acción de HTML5 para el procesamiento de dichos datos, del cual depende una cantidad significativa de sitios web. Concretamente, no cumple con la Aceptación Universal. Creemos que la prioridad número uno para las partes interesadas debería ser la mejora de este estándar. Esto aumentaría la aceptación general, especialmente si los sitios web que están listos para la Aceptación Universal pudieran optar por utilizar un `<input type="email">` e indicar su capacidad para aceptar estas direcciones.



## Integrantes del equipo de investigación

**[Supervisor] Paulo Milliet Roque** tiene una amplia trayectoria en la industria de la tecnología y experiencia en comercio internacional. Ha realizado negociaciones con más de 100 compañías en diversos países. Es cofundador y vicepresidente de la Asociación Brasileira de Empresas de Software (ABES). La ABES fue fundada en 1986 y es la entidad más representativa del sector con aproximadamente dos mil empresas asociadas y afiliadas.

**[Coordinador] Mark W. Datysgeld** es embajador del UASG. Posee una licenciatura y una maestría en Relaciones Internacionales, se especializa en gobernanza de Internet y los impactos de la tecnología en la formulación de políticas públicas y privadas. Mediante la marca Governance Primer, asesora a empresas y particulares en cuanto a su participación en instituciones y eventos internacionales relacionados con la tecnología.

**[Examinador líder] Sávyo Vinicius de Moraes** es embajador y participante del Programa NextGen de la ICANN. Realiza investigaciones en materia de seguridad centrándose en la Internet de las Cosas (IoT) en entornos de hogares y pequeñas oficinas (SOHO) con el objetivo de mitigar el impacto de los ataques masivos a dispositivos conectados a la IoT. Se dedica activamente al desarrollo web y a la administración de sistemas, campos en los que cuenta con experiencia profesional.

**[Examinador] Edson Celio Ferreira Araujo** es un joven estudiante de Ingeniería Informática que se desempeña como desarrollador de sistemas en Grendene S/A. En su tiempo libre, colabora en proyectos de código abierto y trabaja en iniciativas de gobernanza de Internet. Es egresado del Programa IGF para la Juventud.

**[Examinador] Jonas Mendes Fiorini** es un joven técnico en Informática que actualmente estudia Ingeniería Informática en la Universidade Federal do Espírito Santo (UFES). Participa en proyectos relacionados con la inclusión digital y soluciones en red. Es un promotor entusiasta del software libre desde 2012. Es egresado del Programa IGF para la Juventud.

**Estos son nuestros valiosos colaboradores:** Don Hollander, Ajaay Data, Sarmad Hussain, Jennifer Chung, Nivaldo Cleto, Rodrigo de la Parra, Daniel Fink, Seda Akbulut, Vanda Scartezini, Rubens Kuhl y NIC.br.



## Metodología

### Procedimientos

La base de nuestra metodología proviene del documento "UASG017: Evaluación de Sitios Web para la Aceptación de una Variedad de Direcciones de Correo Electrónico", un estudio clave realizado en septiembre de 2017 para registrar por primera vez el cumplimiento de la Aceptación Universal a gran escala. Este estudio, realizado en 2019, profundiza en los componentes esenciales de la investigación anterior con la intención de avanzar en esos métodos.

Una importante decisión metodológica fue no utilizar el conjunto de sitios web de la prueba realizada en 2017. Si bien existe un valor significativo al comparar si los sitios web han progresado o no en su tasa de cumplimiento, el hecho es que no se trataría de una evaluación del estado de los sitios web más importantes del mundo, sino de un estudio totalmente diferente; un estudio que es igualmente válido.

También hubo algunas divergencias en los procedimientos sobre los cuales se basa esta decisión, incluido el hecho de que el conjunto de datos final de la prueba anterior contenía alrededor de 750 entradas, mientras que la versión de 2019 llegó a 1000. Sería poco realista incluir de forma retroactiva más entradas en el conjunto de datos de 2017, por lo que una base de datos nueva era la opción que tenía más sentido en este caso. En adelante, se puede optar por cualquiera de estas dos alternativas.

El primer paso en este estudio fue hacer una lista de los URL comprobables entre los 1000 sitios web más importantes del mundo según la herramienta de análisis competitivo Alexa. Esta acción se concretó a principios de 2019. Los sitios web seleccionados tenían que cumplir con dos criterios: no ser *malware* y tener un campo de entrada de correo electrónico disponible en alguna parte de sus páginas. No se tuvieron en cuenta las cuestiones de contenido, por lo que todos los sitios web fueron tratados como igualmente válidos, independientemente de su temática.

De la lista inicial de los 1000 sitios web más visitados, solo 527 cumplían con los criterios de la prueba, lo que llevó al equipo a seguir analizando la lista para poder encontrar más candidatos. Finalmente, se llegó a los 1922 sitios web que se convirtieron en la selección final para el conjunto de datos. Si fuéramos más precisos a la hora de nombrar el conjunto de datos, diríamos que se trata de la lista de los "1000 sitios web más comprobables según Alexa".

El equipo también consultó la nacionalidad de los sitios web mediante un registro WHOIS que todavía estaba actualizado en ese momento para tener una perspectiva del alcance geográfico de la prueba. Esto puso de relieve el hecho de que gran parte de la lista sigue concentrada en el hemisferio occidental, particularmente dominada por Estados Unidos. En el hemisferio oriental, China, Rusia, Japón, India, Corea del Sur, Taiwán y Hong Kong se destacan por tener la presencia más importante.



A continuación, se presenta el desglose completo por código de país.

Cabe aclarar que la suma no llega a 1000 debido a la imposibilidad de detectar el origen de una pequeña cantidad de sitios web.

<b>AE</b>	3	<b>CN</b>	47	<b>HK</b>	8	<b>LV</b>	1	<b>AS</b>	2	<b>VC</b>	1
<b>AM</b>	1	<b>CR</b>	1	<b>ID</b>	7	<b>MA</b>	2	<b>SC</b>	1	<b>VE</b>	1
<b>AR</b>	5	<b>CY</b>	8	<b>IE</b>	3	<b>MU</b>	1	<b>SE</b>	3	<b>VG</b>	4
<b>AT</b>	1	<b>CZ</b>	11	<b>IL</b>	6	<b>MX</b>	2	<b>SG</b>	2		
<b>AU</b>	13	<b>DE</b>	29	<b>IM</b>	1	<b>NG</b>	1	<b>SI</b>	1		
<b>BA</b>	1	<b>DK</b>	3	<b>IN</b>	16	<b>NL</b>	4	<b>SK</b>	1		
<b>BD</b>	1	<b>DO</b>	1	<b>IR</b>	7	<b>NO</b>	2	<b>TH</b>	2		
<b>BE</b>	1	<b>EC</b>	1	<b>IT</b>	14	<b>PA</b>	35	<b>TN</b>	1		
<b>BG</b>	1	<b>EG</b>	2	<b>JP</b>	24	<b>PE</b>	1	<b>TO</b>	3		
<b>BR</b>	13	<b>ES</b>	18	<b>KE</b>	1	<b>PH</b>	3	<b>TR</b>	9		
<b>BS</b>	10	<b>FR</b>	26	<b>KR</b>	11	<b>PL</b>	10	<b>TW</b>	12		
<b>BY</b>	1	<b>GB</b>	9	<b>KZ</b>	2	<b>PT</b>	4	<b>UA</b>	2		
<b>CA</b>	30	<b>GI</b>	1	<b>LA</b>	1	<b>RO</b>	2	<b>UK</b>	23		
<b>CH</b>	2	<b>GR</b>	3	<b>LU</b>	13	<b>RU</b>	33	<b>US</b>	381		

A continuación, se creó un conjunto de seis buzones de correo para interactuar con los campos de formularios de los URL seleccionados, con un incremento progresivo en sus niveles de complejidad en relación con los dominios legados estándar que utilizan el código ASCII. Las mismas direcciones se mantuvieron desde el inicio hasta el final del estudio. La lista de buzones de correo es la siguiente:

<b>Caso de prueba</b>	<b>Ejemplo</b>
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

Hubo un cambio significativo con respecto a la prueba de 2017, dado que se descartó el buzón de correo "ascii@ascii.idn". Esta no fue la intención inicial del equipo, puesto que se supone que es un caso de uso admitido según las RFC relevantes. Sin embargo, tras intentar registrar un correo electrónico de caso de prueba utilizando ese formato en cinco registros diferentes, los cuales resultaron en errores debido a que el formato no era compatible, llegamos a la conclusión, junto con los líderes del UASG, de no incluirlo en la prueba actual ya que no se lo está adoptando de manera significativa en este momento.



Un elemento heredado de la prueba anterior fue el caso de prueba "arabic.arabic@arabic". Si se lo representa utilizando su nomenclatura adecuada, sería el caso de prueba de derecha a izquierda (RTL). Hemos corregido esta denominación errónea en cierta medida al agregarle "RTL" a la etiqueta, pero la falta de pruebas en hebreo, persa, urdu, sindi, yidis y otros idiomas relevantes nos impidió hacer una transición completa del nombre. Esta situación debe tenerse en cuenta en pruebas adicionales.

El siguiente paso en el procedimiento fue la prueba práctica de los formularios de correo electrónico, insertando la dirección de los casos de prueba uno por uno y enviando los formularios para que fueran procesados por los sitios web. Algunos formularios se omitieron en esta etapa debido a los requisitos de verificación de SMS o captchas irresolubles en un teclado estándar occidental. En este segundo caso, aunque la cifra no fue lo suficientemente elevada como para arrojar resultados, existe la posibilidad de que haya tenido cierta influencia en el estudio.

Cada vez que este proceso se llevó a cabo con éxito, al sitio web se le asignaron seis calificaciones de "aceptado" o "rechazado" según los siguientes criterios:

Se marcaba como "aceptado" en los siguientes casos:

- ✓ La presentación daba como resultado un mensaje de éxito.
- ✓ La presentación se aceptaba y no se informaba ningún error.
- ✓ Se visualizaba un mensaje de correo electrónico "ya registrado"<sup>3</sup>.

Se marcaba como "rechazado" en los siguientes casos:

- ✗ El sitio web mostraba un error tras introducir la dirección.
- ✗ La presentación devolvía un mensaje de error.
- ✗ No se permitía realizar la presentación.

La tabla resultante tenía el siguiente formato:

Sitio web	Correo electrónico 1	Correo electrónico 2	Correo electrónico 3	Correo electrónico 4	Correo electrónico 5	Correo electrónico 6
test.org	Aceptado	Aceptado	Aceptado	Rechazado	Rechazado	Rechazado
ページ. 日本	Aceptado	Aceptado	Aceptado	Aceptado	Aceptado	Rechazado

Al realizar la prueba de los primeros 100 sitios web, el equipo tardó un promedio de 10 minutos con cada uno de ellos. A medida que avanzaba la prueba, el tiempo promedio se redujo a aproximadamente cinco minutos por sitio web, lo cual consideramos una expectativa realista para futuros intentos si alguien repite esta metodología.

A medida que se realizaban las pruebas, el código HTML de cada página se guardaba de forma local para poder extraer y analizar los códigos de validación del campo de correo electrónico. Este proceso insumió un tiempo variable en función del tipo de solución y tecnología empleados. Si la validación se realizaba en el lado del servidor o en HTML5, esto podía identificarse en un minuto o menos. Sin embargo, otras tecnologías como JavaScript

<sup>3</sup>En este caso, el sitio web tenía una base de datos compartida con otra que ya estaba registrada.





podían tardar hasta 15 minutos por sitio web, sobre todo porque el código a menudo estaba minificado<sup>4</sup>.

En lugar de simplemente asignar estos resultados en una hoja de cálculo digital, el equipo optó por organizarlos en una base de datos de MongoDB, una plataforma orientada a JSON que no necesita ser modelada. Esta solución resultó ser ampliable y ligera. Además, todos los examinadores tuvieron acceso a la versión más actualizada de las pruebas en todo momento. Fue fácil visualizar los datos e incluso un usuario menos experimentado podría haber realizado la operación. En general, consideramos que es una opción superior a una base de datos SQL o una hoja de cálculo compartida en línea.

Una vez que se crea una base de datos maestra en MongoDB, se puede hacer un número ilimitado de "recopilaciones", las cuales son análogas a las tablas. Las recopilaciones contienen todos los resultados pertinentes a una prueba en particular y, una vez nombradas correctamente, existen como conjuntos de datos independientes que son interoperables. Al adoptar la nomenclatura "ua\_<scope>\_<year>", todos podrían recopilarse en la misma base de datos que se mantiene en nombres legibles para el ser humano.

Por ejemplo, la recopilación para este estudio se llama "ua\_global\_2019". Si la del año siguiente se llama "ua\_global\_2020", podrían coexistir sin problemas y los investigadores podrían hacer referencias cruzadas de sus datos. Es importante tener en cuenta que los nombres de las recopilaciones distinguen entre mayúsculas y minúsculas. Un futuro estudio que se lleve a cabo en México podría llamarse "ua\_regional\_Mexico\_2019", y así sucesivamente.

El paso final fue agregar todos los datos y convertir la base de datos resultante a un archivo CSV para su procesamiento y evaluación a fin de elaborar este informe. Cualquier CSV que siga el esquema de metadatos detallado a continuación puede ser importado a la base de datos con facilidad con el siguiente comando, el cual supone un entorno de Ubuntu:

```
mongoimport --db ua_database --collection <scope>_<year> --  
host=<hostname> --username=<username> --password=<password> --  
drop --type CSV --file <file_address>/ua_global_2019.csv --  
headerline
```

## Esquema de metadatos

El equipo eligió MongoDB para el almacenamiento de los datos y creó un esquema de metadatos que esperamos se emplee en futuras pruebas, de modo que los resultados de diferentes muestreos puedan compartir el mismo estándar y ser comparados cuando se desee según la perspectiva que los investigadores elijan.

A continuación, presentamos y explicamos este estándar.

---

<sup>4</sup> Definición: eliminar datos innecesarios o redundantes sin afectar la forma en que el navegador procesa el recurso, como el formato y los comentarios del código; eliminar el código no utilizado; utilizar nombres de variables y funciones más cortos, etc. [Wikipedia].



```
{
  "_id": {"$oid": "00001"}
  "domain": "test.org",
  "url":
  "https://www.test.org/signup",
  "rank": "1000",
  "testable": "Yes",
  "code": "<input
type='email'>",
  "comments": "Field triggers
a captcha",
  "mailboxes":
  {
    "mail1": "Accepted",
    "mail2": "Accepted",
    "mail3": "Accepted",
    "mail4": "Rejected",
    "mail5": "Rejected",
    "mail6": "Rejected"
  }
}
```

Estas son las funcionalidades de cada cadena de caracteres:

- } **\_id**: identificador único autogenerated dentro de toda la base de datos.
- } **domain**: dominio, no precedido por "www".
- } **url**: URL que contiene la ruta completa a la página que contiene el formulario.
- } **rank**: posición del sitio web dentro de esa recopilación en particular.
- } **testable**: indica si el equipo finalmente logró probar el sitio web.
- } **code**: contiene la cadena de caracteres que valida el formulario, si se la encuentra.
- } **comments**: cualquier comentario pertinente.
- } **mailboxes**: lista de calificaciones para cada caso de prueba.

Para pruebas futuras, contemplamos la inclusión de un campo "**EAI**", que indicaría si el servidor de correo de un dominio determinado admite la Internacionalización de Direcciones de Correo Electrónico, lo que significa fundamentalmente que podrían intercambiar correo en UTF-8 y, por consiguiente, en Unicode. Esto nos daría una idea más clara de cuál es el nivel general de compatibilidad para esta tecnología, teniendo en cuenta cuántos puntos de falla existen en el camino hacia el intercambio exitoso de mensajes desde direcciones más complejas.

Para este informe concreto de 2019, se había incluido una cadena de caracteres "**country**" (país) para evaluar mejor la procedencia geográfica de los elementos de la lista. Sin embargo, esto no se sugiere como parte del estándar debido a los cambios realizados en la base de datos de WHOIS que impiden búsquedas masivas como la que realizó el equipo, lo cual haría que fuera poco factible de aquí en adelante. Si el RDAP finalmente permite que se realicen dichas consultas, la cadena de caracteres podría volver a ser viable.

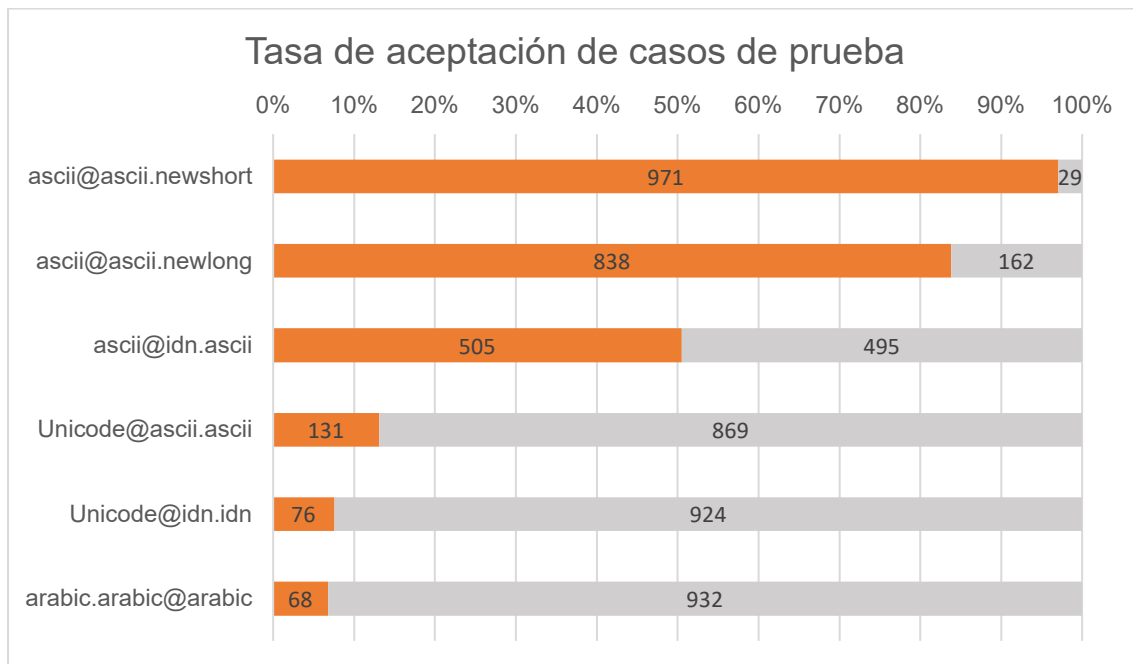


## Resultados

### Totales de pruebas en cifras

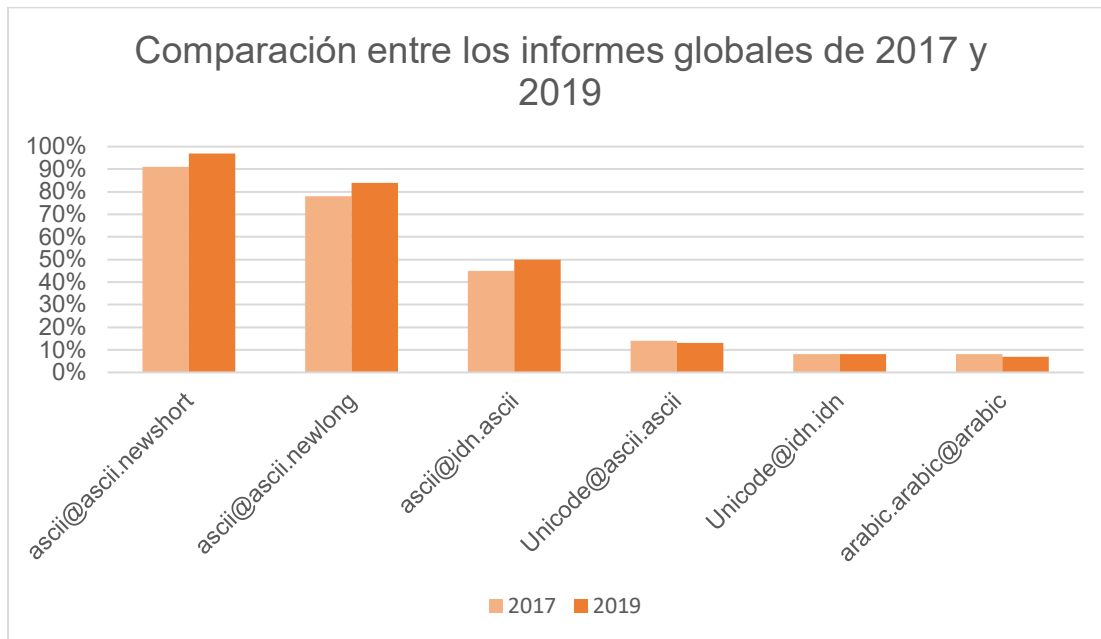
Caso de prueba	Aceptado	Rechazado
ascii@ascii.newshort	971	29
ascii@ascii.newlong	838	162
ascii@idn.ascii	505	495
Unicode@ascii.ascii	131	869
Unicode@idn.idn	76	924
arabic.arabic@arabic (RTL)	68	932

### Gráfico con los totales de pruebas





## Comparación entre totales de pruebas de 2017 y 2019





## Evaluación y correlaciones

En esta sección repasaremos algunos de los resultados que el equipo obtuvo durante la evaluación de los códigos y pondremos de manifiesto cuáles son las malas prácticas en cuanto al uso de los formularios de correo electrónico en la web. También tuvimos la oportunidad de llevar a cabo un experimento a pequeña escala aprovechando la recopilación de datos regionales.

Una consideración importante es que, al igual que en el estudio de 2017, no existe un enfoque unificado para la codificación de las funciones de validación, salvo en el caso de HTML5 y algunos códigos de escritura estandarizados, como "jquery-validation.js". Una cantidad significativa de sitios web depende de Expresiones Regulares (RegEx) en JavaScript y logra diversos grados de éxito en su uso. Sin embargo, por regla general, muy pocas veces son una buena solución.

### Errores de validación de la muestra

Los sitios web con un mejor código en general, pero con un cumplimiento deficiente de la Aceptación Universal, rechazaron el correo electrónico directamente al introducirlo en el formulario. Los sitios web con un código más deficiente aceptaron la dirección a primera vista pero devolvieron un mensaje de error. En algunos casos, el sitio web identificó los casos de prueba más complicados como maliciosos y los clasificó como ataques o intentos de piratería informática.

Estos son algunos de los mensajes más frecuentes (en sus distintas versiones) ante el rechazo de un caso de prueba:

Ingrese una dirección de correo electrónico válida.
Formato de correo electrónico inválido.
Lo sentimos, ese correo electrónico no se ve bien. Compruebe que sea un correo electrónico válido.
El texto antes del carácter "@" no debería contener símbolos.
Se produjo un error.

Se destaca un tipo de mensaje en particular: hay un problema si el carácter "@" está precedido por símbolos. Esto fue bastante uniforme en nuestras observaciones y, al parecer, es la principal idea errónea que está causando problemas en términos de validación.

### Rechazo de todas las direcciones de correo electrónico

El equipo encontró algunos ejemplos particularmente malos de prácticas de codificación entre los 1000 mejores sitios web del mundo. A continuación, separamos un ejemplo en JavaScript que generó error en cada caso de prueba, proveniente nada menos que de un sitio web vietnamita. Además, no se trata de una muestra aislada. Otros sitios web de la lista cometen el mismo error de diversas maneras.

RegEx:

```
/^[[-a-z0-9~!$%^&*_=+}{\'?]+(\.[-a-z0-9~!$%^&*_=+}{\'?]+)*@([-a-z0-9_][-a-z0-9_]*\.[-a-z09_]+)*\.(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
```



Este fragmento de código hace una comprobación rigurosa de parámetros muy específicos que definen la estructura de una dirección de correo electrónico legada, con especial atención a la inclusión de algunos de los TLD patrocinados en una “lista de TLD aprobados”, así como a la inclusión redundante de los TLD legados, como “.edu” y “.org”, en la misma lista. No tiene mucho sentido, pero se encontraron variaciones de esta situación en otras instancias.

### Pruebas de correlación regional

Aunque no era uno de los principales objetivos del estudio, al equipo le interesó tratar de correlacionar el cumplimiento de la Aceptación Universal con los principales sitios web de países que emplean códigos de escritura diferentes del latino, con la expectativa de que sus niveles de cumplimiento podrían ser más altos. Tras estudiar los códigos de escritura cirílico y han, concluimos que **esta correlación no parece existir** a este nivel. Quizás, en futuras pruebas locales sea posible evaluar mejor los 100 sitios web más importantes a nivel regional y determinar si tienen mejores resultados, lo cual es una cuestión importante de cara al futuro.

Los resultados pueden observarse a continuación:

- **Caso del código de escritura han:** el grupo incluyó 47 sitios web de China, 8 de Hong Kong, 2 de Singapur y 12 de Taiwán; en total, solo 4 sitios web de China pudieron procesar el caso de prueba moderado "unicode@ascii.ascii".
- **Caso del código de escritura cirílico:** el grupo incluyó 37 sitios web de Rusia y 2 de Ucrania; en total, solo 5 sitios web de Rusia pudieron procesar el caso de prueba moderado "unicode@ascii.ascii".



## HTML5: el obstáculo para la Aceptación Universal

En "ua\_regional\_Brazil\_2018"<sup>5</sup>, los resultados indicaron que el 30 % de los 50 sitios web más importantes del país utilizaron `<input type="email">` como su solución para la validación de direcciones de correo electrónico por parte del navegador. Tal como lo esperaba, el equipo halló un patrón similar en el estudio global. La tasa de uso global de esta cadena de caracteres oscila entre el 20 y el 30 % de acuerdo con los patrones que hemos observado.

El problema es que, en pleno 2019, el estándar HTML5 aún no se ha puesto al día con la Aceptación Universal. Esto nos deja preocupados, dado que el despliegue del estándar continúa creciendo y los desarrolladores confían cada vez más en sus funcionalidades para que su código funcione en tantos escenarios como el estado actual de la web lo requiera.

Este es el patrón actual de aceptación para `<input type="email">`:

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

Este resultado refuerza nuestro postulado anterior: al parecer, la principal idea errónea es que no debería haber símbolos antes del carácter "@".

Si entre el 20 y el 30 % de los sitios web utilizan esta cadena de caracteres, entonces, en sentido inverso, esto significa que el mismo número de sitios web está generando errores en esta etapa de validación debido a la falta de estándares adecuados en HTML5. En otras palabras, en lo que respecta a los navegadores, esta es la acción más significativa que se puede tomar para promover la Aceptación Universal en este momento.

Después de investigar el estado de este asunto, encontramos que la versión preliminar 5.3 de la especificación, en la sección relacionada con la validación por correo electrónico<sup>6</sup>, prevé el cumplimiento de la RFC 6531 y la RFC 5890; con lo cual, el estándar debe cumplir con la Aceptación Universal. Sin embargo, esta revisión tiene el estado de "primera versión preliminar" desde 2017, sin fecha prevista para que se convierta en una "recomendación preliminar" o en una "recomendación final".

Consideramos que esto es preocupante dada la importancia del asunto para que la comunidad concrete la Aceptación Universal, particularmente en lo que se refiere a los nombres internacionalizados. Este es un tema que la comunidad no puede dejar pasar. Es

---

<sup>5</sup><http://central.abesssoftware.com.br/Content/UploadedFiles/Arquivos/Estudos,%20Pesquisas%20e%20Pareces/Universal-Acceptance-in-Brazil-2018.pdf>

<sup>6</sup><https://www.w3.org/TR/html53/sec-forms.html#email-state-typeemail>



necesario tomar amplias medidas en nuestra relación con el W3C, el WHATWG y los proveedores de navegadores a fin de acelerar el proceso de implementación de este aspecto de la especificación.

Esto resulta particularmente oportuno, ya que el W3C y el WHATWG recientemente acordaron adoptar un enfoque conjunto para el desarrollo de estándares<sup>7</sup>. Es importante comunicarse con las partes interesadas y exigir que se agregue la opción "eaimail" a los sitios web que están preparados para la Aceptación Universal, así como presionar a los principales proveedores de navegadores para que indiquen su interés en hacer que esta opción sea viable.

El equipo de investigación recomienda que se le asigne una prioridad alta a la resolución de este problema.

---

<sup>7</sup> <https://www.w3.org/blog/2019/05/w3c-and-whatwg-to-work-together-to-advance-the-open-web-platform/>





## ¿Pueden automatizarse estos informes?

Durante la elaboración de este estudio, el UASG planteó preguntas, en las reuniones de la ICANN y durante las sesiones de planificación, acerca de la viabilidad de un software automatizado para la producción de informes como éste. El equipo intentó encontrar una respuesta práctica a esa pregunta, pero llegó a la conclusión de que no es posible automatizar estas pruebas por el momento.

Estas son nuestras principales preocupaciones:

- Hay demasiada variedad en lo que respecta al modo de codificar los sitios web, los códigos de escritura que utilizan y el modo de organizar sus formularios. En general, sería prácticamente imposible representar todas las variantes.
- La minificación se utiliza ampliamente para reducir los tiempos de carga de los sitios web y códigos de escritura. En ocasiones, esto requiere que alguien con experiencia en codificación revise los resultados.
- Una cantidad significativa de sitios web requiere algún tipo de Captcha o tiene algún matiz en sus formularios, lo que hace que la automatización sea inútil.

Algunos de estos argumentos también son válidos en cuanto a la tercerización del estudio a personas que no son expertas, como en el caso de la utilización de Mechanical Turk de Amazon. Si bien eso puede funcionar para llevar a cabo las pruebas reales, es necesario coordinar la preparación de la lista de dominios, la extracción de códigos y la elaboración de un informe significativo que sea más que una estadística. Sin embargo, es mucho más viable que la automatización.



## Conclusión

Si tomamos como referencia los 100 sitios web más importantes del mundo, solo cinco de ellos aceptaron todos los casos de prueba: “quora.com”, “espn.com”, “spotify.com”, “txxx.com” [NSFW] y “godaddy.com”. Suponiendo que omitiéramos los casos de prueba difíciles “unicode@idn.idn” y árabe (RTL), aun así obtendríamos tan solo nueve sitios web que aceptan los primeros cuatro casos de prueba. Debido a que se trata de las páginas más visitadas del mundo, esto es muy preocupante.

Si bien somos optimistas en cuanto a la perspectiva de lograr el pleno cumplimiento a corto plazo en los dominios con extensiones breves (*newshorts*), y nos parece realista que los dominios con extensiones más largas (*newlongs*) lo logren a largo plazo, es evidente que los dominios internacionalizados necesitan un impulso y es necesario emprender acciones proactivas para que prosperen. No solo es necesario desarrollar políticas más sólidas. También es necesario concientizar acerca de estos correos electrónicos y dominios en el ecosistema de Internet, lo cual solo se logrará mediante una difusión proactiva.

En conclusión, creemos que un paso importante sería el desarrollo de un tablero de control que permita la evaluación de todos los informes sobre Aceptación Universal que se generen utilizando el esquema de metadatos propuesto en este documento. Esto generaría un repositorio duradero que permitiría a cualquier parte interesada realizar mediciones realistas y serviría como herramienta de referencia para toda la comunidad.