Universal Acceptance –
Getting ready for the Expansion of the DNS

Lars Steffen
Get Engaged in ICANN Seminar / 21 February 2018
Warm-Up
Warm-up Exercise

According to w3techs, which of the following pie charts most closely represents the fraction of websites on the Internet that are primarily English language based content?

25% English
75% All other

50% English
50% All other

75% English
25% All other

Actual data can be found at https://w3techs.com/technologies/history_overview/content_language/ms/y. Data shown is approximate.
Warm-up Exercise: Solution

According to w3techs, which of the following pie charts most closely represents the fraction of websites on the Internet that are primarily English language based content?

- 25% English, 75% All other
- 50% English, 50% All other
- 75% English, 25% All other

Actual data can be found at https://w3techs.com/technologies/history_overview/content_language/ms/y. Data shown is approximate.
**Warm-up Exercise**

Each of the 3 groups below contain lists of Top Level Domains (TLDs) that are valid (approved and delegated by ICANN), except that each list contains one made-up or invalid TLD. **Which TLD in each group is invalid?**

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
</tr>
</thead>
<tbody>
<tr>
<td>嘉里</td>
<td>ABC</td>
<td>GMAIL</td>
</tr>
<tr>
<td>ANALYTICS</td>
<td>مصر</td>
<td>قطر</td>
</tr>
<tr>
<td>BLOCKBUSTER</td>
<td>ATHLETA</td>
<td>JOY</td>
</tr>
<tr>
<td>DIAMONDS</td>
<td>அணுநிசைக்குத்</td>
<td>LIKE</td>
</tr>
<tr>
<td>HOTELES</td>
<td>CANCERRESEARCH</td>
<td>ஹஸ</td>
</tr>
<tr>
<td>广东</td>
<td>CITIC</td>
<td>ONYOURSIDE</td>
</tr>
<tr>
<td>MOVISTAR</td>
<td>新加坡</td>
<td>OOO</td>
</tr>
<tr>
<td>டுதுதிப்பு</td>
<td>ESURANCE</td>
<td>فلسطين</td>
</tr>
<tr>
<td>REALLY</td>
<td>FAKE</td>
<td>SILLY</td>
</tr>
<tr>
<td>政务</td>
<td>ไทย</td>
<td>SUCKS</td>
</tr>
</tbody>
</table>
Warm-up Exercise

The 3 invalid TLDs are highlighted below in red. There are 1541 valid TLDs (as of January 2\textsuperscript{nd}, 2018), and the list is continuing to grow.

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
</tr>
</thead>
<tbody>
<tr>
<td>嘉里</td>
<td>ABC</td>
<td>GMAIL</td>
</tr>
<tr>
<td>ANALYTICS</td>
<td>مصر</td>
<td>قطر</td>
</tr>
<tr>
<td>BLOCKBUSTER</td>
<td>ATHLETA</td>
<td>JOY</td>
</tr>
<tr>
<td>DIAMONDS</td>
<td>งินจักรีแสดง</td>
<td>LIKE</td>
</tr>
<tr>
<td>HOTELES</td>
<td>CANCERRESEARCH</td>
<td>hహసు</td>
</tr>
<tr>
<td>广东</td>
<td>CITIC</td>
<td>ONYOURSIDE</td>
</tr>
<tr>
<td>MOVISTAR</td>
<td>新加坡</td>
<td>OOO</td>
</tr>
<tr>
<td>禮拜</td>
<td>ESURANCE</td>
<td>فلسطين</td>
</tr>
<tr>
<td>REALLY</td>
<td>FAKE</td>
<td>SILLY</td>
</tr>
<tr>
<td>政务</td>
<td>ไทย</td>
<td>SUCKS</td>
</tr>
</tbody>
</table>

Group B: งินจักรีแสดง, CANCERRESEARCH, CITIC, 新加坡, ESURANCE, FAKE, ไทย

Group C: GMAIL, قطر, JOY, LIKE, hహసు, ONYOURSIDE, OOO, فلسطين, SILLY, SUCKS
Building Blocks: **gTLDs**

* Starting in 2013, ICANN (the organization responsible for the creation and maintenance of TLD assignments) approved the creation of a large number of new TLDs. These new TLDs can represent brands, communities of interest, geographic communities (cities, regions) and more generic concepts. Collectively, all of these new TLDs are known as Generic Top Level Domains (gTLDs).

<table>
<thead>
<tr>
<th>Original Seven gTLDs</th>
<th>Common cc TLDs</th>
<th>New gTLDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>.com</td>
<td>.de</td>
<td>.top</td>
</tr>
<tr>
<td>.org</td>
<td>.cn</td>
<td>.xyz</td>
</tr>
<tr>
<td>.gov</td>
<td>.uk</td>
<td>.loan</td>
</tr>
<tr>
<td>.edu</td>
<td>.nl</td>
<td>.club</td>
</tr>
<tr>
<td>.mil</td>
<td>.eu</td>
<td>.网址</td>
</tr>
<tr>
<td>.net</td>
<td>.ru</td>
<td>.信息</td>
</tr>
<tr>
<td>.int</td>
<td>.tk</td>
<td>.コム</td>
</tr>
</tbody>
</table>
The Reality
Hi!
Hi!

Warning

"Борис@пример.рф" does not appear to be a valid email address. Verify the address and try again.

[Options: Cancel, Send Anyway]
Definition

*Universal Acceptance* (UA) ensures that all **domain names** and **email addresses** can be used by all Internet-enabled **applications**, **devices** and **systems**.
Universal Acceptance Steering Group
Who is involved...
UASG Activities

Review: Popular Websites, Dev Frameworks, Browsers, OS
Build: Use Cases, Test Environments, EAI Community
Outreach: Live Workshops, Panel Discussions, Presentations
Writing: Knowledge Databases, Whitepapers, Quick Guides
Why should you care?
Email:
Your email...

Password:
Your password...
Internet

إنترنت
How to get Universal Acceptance ready?
Universal Acceptance ready

* Accept, Support, Use and Show Unicode.
* Store Information in UTF-8.
* Accept Strings up to 256 characters.
* Validate as little as possible.
  * Verify TLDs against authoritative table.
  * Query domain name against DNS.
Universal Acceptance ready

* When developing app / service, or operating a registry, consider languages supported.

* Perform code reviews.

* Check code points that were not defined when application / service was created.
Universal Acceptance ready

* Consider mixed-script addresses become more common.

* Convert non-Unicode to Unicode before display.

* Use Unicode IDNA Compatibility Processing to match user expectations.

* End user should see “everyone.みんな” vs. “everyone.xn--q9jyb4c.”
Some technical information
Building Blocks: **ASCII and Unicode**

A character mapping associates characters with specific numbers. Many different code pages have been created over time for different purposes, but for this topic we will focus on only two: **ASCII and Unicode**.

**ASCII**

Most of the text currently displayed on the internet is in the Latin character set. This character set is included in the American Standard Code for Information Interchange (ASCII, or US-ASCII) character-encoding scheme. ASCII is an older encoding scheme and was based on the English language. For historical reasons, it became the standard character encoding scheme on the Internet.

**Unicode**

Because most writing systems do not use the Latin character set, alternate encodings have also been adopted. The most common form of Unicode is called Universal Coded Character Set Transform Format 8-bit (UTF-8).

To see all Unicode character code charts, go to: [http://unicode.org/charts](http://unicode.org/charts)
Building Blocks: ASCII and Unicode (cont.)

**ASCII**

ASCII uses only 7 bits per character, which limits the set to 128 characters, not all of which can be used in domain names.

Domain names are limited to the characters A-Z, the numbers 0-9, and hyphen “-”.

**Unicode**

Unicode, also known as the Universal Coded Character Set (UCS), is capable of encoding more than 1 million characters.

Each of these Unicode characters is called a code point.
Building Blocks – Internationalized Domain Names and Email Addresses

* The use of Unicode enables domain names and email addresses to contain non-ASCII characters.

  * Domain names that use non-ASCII characters are called Internationalized Domain Names (IDNs). The internationalized portion of a domain name can be in any level – not just the TLD but also the other labels.

  * Email addresses that use non-ASCII characters are called Internationalized Email Addresses. The internationalized portion can be at the local or domain part of the address.

* Since the DNS itself previously only used ASCII, it was necessary to create an additional encoding to allow non-ASCII Unicode code points to be converted into ASCII strings, and vice versa.
### EAI, examples

<table>
<thead>
<tr>
<th>username</th>
<th>domain</th>
<th>TLD</th>
<th>username</th>
<th>domain</th>
<th>TLD</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:user@example.app">user@example.app</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>app.مثال@مستخدم</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Left to Right (LTR) Scripts

- Username: `user`
- Domain: `example.app`

### Right to Left (RTL) Scripts

- Username: `المستخدم.مثال`
- Domain: `app`

### More Examples of (imaginary) Email Addresses including IDNs

- `user@example.みんな`(Uses internationalized TLD)
- `user@大坂.info`(Uses internationalized 2nd level domain)
- `用戶@example.lawyer`(Uses internationalized user name and new gTLD)
POP & IMAP Servers

* Post Office Protocol version 3 (POP3) supports international strings encoded in UTF-8 in usernames, passwords, mail addresses, message headers, and protocol-level text strings (see RFC 6856).

* Internet Message Access Protocol (IMAP) supports UTF-8 encoded international characters in user names, mail addresses, and message headers (see RFC 6855).
Items for Email Service Providers to Consider

* Don’t enforce case-sensitivity of local-part mailbox names.
* Allow the user to enter the email address in any combination of upper- and-lowercase characters so long as the script is correct.
* Don’t issue mailbox names which will duplicate other mailbox names which have the same characters but different cases (e.g. “user@example.TLD”) and “uSer@example.tld”).

Note non-ASCII case folding may not work because it is language specific.
Items for Email Service Providers to Consider

* Offer an all-ASCII mailbox name to the user when they are issued an EAI-compatible mailbox name.
  * If both names alias to the same mailbox (i.e. can be used interchangeably) users will find it easier to initially share addresses with other users who use a different script.
  * Once the ASCII address is initially shared, a user can decide whether to also add the EAI-compatible address to their address book.
* Offer mailbox names which conform to the domain name Label Generation Rules (LGR) for the selected script.
  * Such names are guaranteed to be compatible with the Punycode algorithm.
  * These email addresses can easily be shared by users with their friends and colleagues who do not use their same writing method; the colleague or friend can address email to such an address, or create an address book entry, using the A-label format.
  * Upon use, the client MUA software should convert the A-Label to the appropriate U-Label, at which point the friend or colleague will possess the EAI formatted email address despite not having a keyboard or IME which supports the target script.
Encountering UA issues?

uasg.tech/global-support-centre
Further information

* Visit uasg.tech
* Visit bit.ly/icannUA
* Visit github.com/icann

* Get your Universal Acceptance Quick Guide!
Thank you!