Introduction to Internationalized Domain Names (IDNs) and Universal Acceptance of Domain Names and Email Addresses (UA)

Sarmad Hussain
Sr. Director IDN and UA Programs

Budapest University
29 April 2021
Forming ASCII Labels

1. Use only Letters
   - Letters [a..z]
   - Label length = 63
   - Other constraints (e.g. on hyphen)

2. Use LDH
   - Letters [a..z]
   - Digits [0..9]
   - Hyphen (-)
   - Label length = 63
## Domain Name Mnemonics in ASCII

Using LDH
- Letters [a..z]
- Digits [0..9]
- Hyphen (-)

```markdown
<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NUL</td>
<td>DLE</td>
<td>space</td>
<td>0</td>
<td>@</td>
<td>P</td>
</tr>
<tr>
<td>1</td>
<td>SOH</td>
<td>DC1</td>
<td>XON</td>
<td>1</td>
<td>A</td>
<td>Q</td>
</tr>
<tr>
<td>2</td>
<td>STX</td>
<td>DC2</td>
<td>&quot;</td>
<td>2</td>
<td>B</td>
<td>R</td>
</tr>
<tr>
<td>3</td>
<td>ETX</td>
<td>DC3</td>
<td>XOFF</td>
<td>3</td>
<td>C</td>
<td>S</td>
</tr>
<tr>
<td>4</td>
<td>EOT</td>
<td>DC4</td>
<td>$</td>
<td>4</td>
<td>D</td>
<td>T</td>
</tr>
<tr>
<td>5</td>
<td>ENQ</td>
<td>NAK</td>
<td>%</td>
<td>5</td>
<td>E</td>
<td>U</td>
</tr>
<tr>
<td>6</td>
<td>ACK</td>
<td>SYN</td>
<td>&amp;</td>
<td>6</td>
<td>F</td>
<td>V</td>
</tr>
<tr>
<td>7</td>
<td>BEL</td>
<td>ETB</td>
<td>'</td>
<td>7</td>
<td>G</td>
<td>W</td>
</tr>
<tr>
<td>8</td>
<td>BS</td>
<td>CAN</td>
<td>(</td>
<td>8</td>
<td>H</td>
<td>X</td>
</tr>
<tr>
<td>9</td>
<td>HT</td>
<td>EM</td>
<td>)</td>
<td>9</td>
<td>I</td>
<td>Y</td>
</tr>
<tr>
<td>A</td>
<td>LF</td>
<td>SUB</td>
<td>*</td>
<td>:</td>
<td>J</td>
<td>Z</td>
</tr>
<tr>
<td>B</td>
<td>VT</td>
<td>ESC</td>
<td>+</td>
<td>:</td>
<td>K</td>
<td>[</td>
</tr>
<tr>
<td>C</td>
<td>FF</td>
<td>FS</td>
<td>,</td>
<td>&lt;</td>
<td>L</td>
<td>\</td>
</tr>
<tr>
<td>D</td>
<td>CR</td>
<td>GS</td>
<td>-</td>
<td>=</td>
<td>M</td>
<td>]</td>
</tr>
<tr>
<td>E</td>
<td>SO</td>
<td>RS</td>
<td>&gt;</td>
<td>N</td>
<td>^</td>
<td>n</td>
</tr>
<tr>
<td>F</td>
<td>SI</td>
<td>US</td>
<td>/</td>
<td>?</td>
<td>O</td>
<td>_</td>
</tr>
</tbody>
</table>
```
Using Letters only
- Letters [a..z]
- Digits [0..9]
- Hyphen (-)

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NUL DLE space</td>
<td>@</td>
<td>P</td>
<td>`</td>
<td>p</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>SOH DC1 XON !</td>
<td>1</td>
<td>A</td>
<td>Q</td>
<td>a</td>
<td>q</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>STX DC2 &quot;</td>
<td>2</td>
<td>B</td>
<td>R</td>
<td>b</td>
<td>r</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ETX DC3 XOFF #</td>
<td>3</td>
<td>C</td>
<td>S</td>
<td>c</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>EOT DC4 $</td>
<td>4</td>
<td>D</td>
<td>T</td>
<td>d</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>ENQ NAK %</td>
<td>5</td>
<td>E</td>
<td>U</td>
<td>e</td>
<td>u</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ACK SYN &amp;</td>
<td>6</td>
<td>F</td>
<td>V</td>
<td>f</td>
<td>v</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>BEL ETB '</td>
<td>7</td>
<td>G</td>
<td>W</td>
<td>g</td>
<td>w</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>BS CAN (</td>
<td>8</td>
<td>H</td>
<td>X</td>
<td>h</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>HT EM )</td>
<td>9</td>
<td>I</td>
<td>Y</td>
<td>i</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>LF SUB *</td>
<td>:</td>
<td>J</td>
<td>Z</td>
<td>j</td>
<td>z</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>VT ESC +</td>
<td>;</td>
<td>K</td>
<td>[</td>
<td>k</td>
<td>{</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>FF FS .</td>
<td></td>
<td>L</td>
<td>\</td>
<td>l</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>CR GS -</td>
<td>=</td>
<td>M</td>
<td>]</td>
<td>m</td>
<td>}</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>SO RS &gt;</td>
<td></td>
<td>N</td>
<td>^</td>
<td>n</td>
<td>~</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>SI US /</td>
<td>?</td>
<td>O</td>
<td>_</td>
<td>o</td>
<td>del</td>
<td></td>
</tr>
</tbody>
</table>
Internationalized Domain Names

Syntax of IDN Labels
Valid U-Label: Unicode code points as constrained by the “LDH” scheme within IDNA 2008

Syntax of IDN Labels
Valid U-Label, further constrained by the “letter” principle for TLDs
### IDN Mnemonics for All Actively Used Scripts

<table>
<thead>
<tr>
<th>060</th>
<th>061</th>
<th>062</th>
<th>063</th>
<th>064</th>
<th>065</th>
<th>066</th>
<th>067</th>
<th>068</th>
<th>069</th>
<th>06A</th>
<th>06B</th>
<th>06C</th>
<th>06D</th>
<th>06E</th>
<th>06F</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>ع</td>
<td>ي</td>
<td>ت</td>
<td>گ</td>
<td>ل</td>
<td>م</td>
<td>ن</td>
<td>ع</td>
<td>م</td>
<td>م</td>
<td>م</td>
<td>م</td>
<td>م</td>
<td>م</td>
<td>م</td>
</tr>
<tr>
<td>1</td>
<td>ي</td>
<td>ك</td>
<td>ه</td>
<td>ك</td>
<td>ك</td>
<td>ك</td>
<td>ك</td>
<td>ك</td>
<td>ك</td>
<td>ك</td>
<td>ك</td>
<td>ك</td>
<td>ك</td>
<td>ك</td>
<td>ك</td>
</tr>
<tr>
<td>2</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>ل</td>
<td>ك</td>
<td>ل</td>
<td>ك</td>
<td>ل</td>
<td>ك</td>
<td>ل</td>
<td>ك</td>
<td>ل</td>
<td>ك</td>
<td>ل</td>
<td>ك</td>
</tr>
<tr>
<td>3</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
</tr>
<tr>
<td>4</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
</tr>
<tr>
<td>5</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
</tr>
<tr>
<td>6</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
</tr>
<tr>
<td>7</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
</tr>
<tr>
<td>8</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
<td>م</td>
<td>ك</td>
</tr>
<tr>
<td>060</td>
<td>061</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>060</td>
<td>061</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>121</td>
<td>122</td>
<td>123</td>
<td>124</td>
<td>125</td>
<td>126</td>
<td>127</td>
<td>128</td>
<td>129</td>
<td>12A</td>
<td>12B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1200</td>
<td>1210</td>
<td>1220</td>
<td>1230</td>
<td>1240</td>
<td>1250</td>
<td>1260</td>
<td>1270</td>
<td>1280</td>
<td>1290</td>
<td>12A0</td>
<td>12B0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1210</td>
<td>1211</td>
<td>1221</td>
<td>1231</td>
<td>1241</td>
<td>1251</td>
<td>1261</td>
<td>1271</td>
<td>1281</td>
<td>1291</td>
<td>12A1</td>
<td>12B1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1220</td>
<td>1221</td>
<td>1232</td>
<td>1242</td>
<td>1252</td>
<td>1262</td>
<td>1272</td>
<td>1282</td>
<td>1292</td>
<td>12A2</td>
<td>12B2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1230</td>
<td>1231</td>
<td>1233</td>
<td>1243</td>
<td>1253</td>
<td>1263</td>
<td>1273</td>
<td>1283</td>
<td>1293</td>
<td>12A3</td>
<td>12B3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1240</td>
<td>1241</td>
<td>1244</td>
<td>1254</td>
<td>1264</td>
<td>1274</td>
<td>1284</td>
<td>1294</td>
<td>12A4</td>
<td>12B4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1250</td>
<td>1251</td>
<td>1255</td>
<td>1265</td>
<td>1275</td>
<td>1285</td>
<td>1295</td>
<td>12A5</td>
<td>12B5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1260</td>
<td>1261</td>
<td>1266</td>
<td>1276</td>
<td>1286</td>
<td>1296</td>
<td>12A6</td>
<td>12B6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1270</td>
<td>1271</td>
<td>1277</td>
<td>1287</td>
<td>1297</td>
<td>12A7</td>
<td>12B7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

IDN Mnemonics for All Actively Used Scripts
### IDN Mnemonics for All Actively Used Scripts

<table>
<thead>
<tr>
<th>HEX</th>
<th>C</th>
<th>J</th>
<th>K</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>4E50</td>
<td>乐</td>
<td>00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4E51</td>
<td>甘</td>
<td>GE-212F</td>
<td>T4-2227</td>
<td>J4-122C</td>
</tr>
<tr>
<td>4E52</td>
<td>甘</td>
<td>GE-4639</td>
<td>HB1-AE62</td>
<td>T1-4724</td>
</tr>
<tr>
<td>4E53</td>
<td>甘</td>
<td>GE-4949</td>
<td>HB1-AE63</td>
<td>T1-4925</td>
</tr>
<tr>
<td>4E54</td>
<td>乔</td>
<td>GE-4537</td>
<td>T3-3045</td>
<td></td>
</tr>
<tr>
<td>4E55</td>
<td>甘</td>
<td>GE-2130</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4E56</td>
<td>甘</td>
<td>GE-3864</td>
<td>HB1-AE64</td>
<td>T1-4965</td>
</tr>
<tr>
<td>4E57</td>
<td>甘</td>
<td>GE-2131</td>
<td>T3-2922</td>
<td>J3-5258</td>
</tr>
<tr>
<td>4E58</td>
<td>甘</td>
<td>GE-333B</td>
<td>HB1-AE8C</td>
<td>T1-5070</td>
</tr>
<tr>
<td>4E59</td>
<td>甘</td>
<td>GE-3220</td>
<td>HB1-AE81</td>
<td>T1-5253</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HEX</th>
<th>C</th>
<th>J</th>
<th>K</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>4E60</td>
<td>甘</td>
<td>OK-4884</td>
<td>K2-213C</td>
<td></td>
</tr>
<tr>
<td>4E61</td>
<td>甘</td>
<td>OK-4864</td>
<td>K2-213D</td>
<td></td>
</tr>
<tr>
<td>4E62</td>
<td>甘</td>
<td>OK-4929</td>
<td>K2-213E</td>
<td></td>
</tr>
<tr>
<td>4E63</td>
<td>甘</td>
<td>OK-4885</td>
<td>K2-213F</td>
<td></td>
</tr>
<tr>
<td>4E64</td>
<td>甘</td>
<td>OK-3923</td>
<td>K1-503A</td>
<td></td>
</tr>
<tr>
<td>4E65</td>
<td>甘</td>
<td>OK-3924</td>
<td>K1-503B</td>
<td></td>
</tr>
<tr>
<td>4E66</td>
<td>甘</td>
<td>OK-4924</td>
<td>K1-503C</td>
<td></td>
</tr>
<tr>
<td>4E67</td>
<td>甘</td>
<td>OK-4885</td>
<td>K1-503D</td>
<td></td>
</tr>
<tr>
<td>4E68</td>
<td>甘</td>
<td>OK-3923</td>
<td>K1-503E</td>
<td></td>
</tr>
<tr>
<td>4E69</td>
<td>甘</td>
<td>OK-3924</td>
<td>K1-503F</td>
<td></td>
</tr>
<tr>
<td>4E6A</td>
<td>甘</td>
<td>OK-4924</td>
<td>K1-504A</td>
<td></td>
</tr>
<tr>
<td>4E6B</td>
<td>甘</td>
<td>OK-4885</td>
<td>K1-504B</td>
<td></td>
</tr>
<tr>
<td>4E6C</td>
<td>甘</td>
<td>OK-4885</td>
<td>K1-504C</td>
<td></td>
</tr>
<tr>
<td>4E6D</td>
<td>甘</td>
<td>OK-4885</td>
<td>K1-504D</td>
<td></td>
</tr>
<tr>
<td>4E6E</td>
<td>甘</td>
<td>OK-4885</td>
<td>K1-504E</td>
<td></td>
</tr>
<tr>
<td>4E6F</td>
<td>甘</td>
<td>OK-4885</td>
<td>K1-504F</td>
<td></td>
</tr>
</tbody>
</table>
# IDN Mnemonics for All Actively Used Scripts

<table>
<thead>
<tr>
<th>HEX</th>
<th>C</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>4E50</td>
<td>004066</td>
<td>-</td>
</tr>
<tr>
<td>4E51</td>
<td>004166</td>
<td>-</td>
</tr>
<tr>
<td>4E52</td>
<td>004266</td>
<td>-</td>
</tr>
<tr>
<td>4E53</td>
<td>004366</td>
<td>-</td>
</tr>
<tr>
<td>4E54</td>
<td>004466</td>
<td>-</td>
</tr>
<tr>
<td>4E55</td>
<td>004566</td>
<td>-</td>
</tr>
<tr>
<td>4E56</td>
<td>004666</td>
<td>-</td>
</tr>
<tr>
<td>4E57</td>
<td>004766</td>
<td>-</td>
</tr>
<tr>
<td>4E58</td>
<td>004866</td>
<td>-</td>
</tr>
<tr>
<td>4E59</td>
<td>004966</td>
<td>-</td>
</tr>
<tr>
<td>4E5A</td>
<td>004A66</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>090</th>
<th>091</th>
<th>092</th>
<th>093</th>
<th>094</th>
<th>095</th>
<th>096</th>
<th>097</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
IDN Mnemonics for All Actively Used Scripts

ASCII
1 script
63 of 127
allowed
IDN Mnemonics for All Actively Used Scripts

Unicode 11.0
??? of 148 scripts
??? of 137,439 allowed

ASCII
1 script
63 of 127 allowed
epic
epic
epic
epic (0065 0070 0069 0063)
epic (0435 0440 0456 0441)
“Same” or Different Domain Labels?

- Example of within-script variant labels (Arabic script)
  - شبكَتِ (06C3 06A9 0628 0634)
  - شبكَة (0629 06A9 0628 0634)
  - شبكَة (0629 0643 0628 0634)

- Example of within-script variant labels (Simplified Chinese and Traditional Chinese)
  - 名称 (540D 79F0)
  - 名稱 (540D 7A31)

- Example of cross-script variant label (Latin script and Cyrillic script)
  - epic (0065 0070 0069 0063)
  - epic (0435 0440 0456 0441)
Generation Panels
- Generate proposals for script specific LGRs, based on community expertise and linguistic, security and stability requirements

Integration Panel
- Integrates them into common Root Zone LGR while minimizing the risk to Root Zone as a shared resource

Label Generation Rules (LGR)
- Which labels are permissible
- Which variant labels exist
- Are there any more constraints?
Scope of Code Point Repertoires for Domain Labels

1. Unicode
   - 148 scripts
   - 137,439

2. IDNA2008
   - 147 scripts
   - 97,973

3. MSR-4
   - 28 scripts
   - 33,511

4. RZ-LGR-4
   - 20 scripts
   - 21,019 – in progress

Second Level Domain

Top Level Domain
IDN Country Code Top-Level Domains

ccTLDs: 62 for 43*

* Successfully evaluated IDN ccTLDs for total countries and territories (as of May 2020)
A Unicode character is categorized as a letter, digit, mark, punctuation or symbol. Which of these are generally allowed for Internationalized domain names (IDNs)?

1. Only letters.
2. Letters and digits.
3. Letters, digits and marks.
4. Letters, digits and symbols.
5. All categories of characters.
Universal Acceptance of Domain Names and Email Addresses (UA)
Universal Acceptance of Domain Names and Email

Goal
All domain names and email addresses work in all software applications.

Impact
Promote consumer choice, improve competition, and provide broader access to end users.
Categories Affected and UA Readiness

- **Domain Names:**
  - **Newer** top-level domain names: `example.sky`
  - **Longer** top-level domain names: `example.international`
  - **Internationalized** domain names: คณ.ไทย

- **Internationalized email addresses (EAI):**
  - **ASCII@ASCII** (new and long TLD) `ekrem@misal.berlin`
  - **ASCII@IDN** `marc@société.org`
  - **Unicode@ASCII** 测试@example.com
  - **Unicode@IDN** пример@тестовая-зона.рф
  - **Unicode@IDN; right to left scripts** ای-میل@مثال.موقع
Scope of UA Readiness

1. Support All Domain Names

   - Accept
   - Validate
   - Process
   - Store
   - Display

2. Support All Email Addresses

   - Accept
   - Validate
   - Process
   - Store
   - Display
Acceptance of Email Addresses by Websites Globally

For details, see UASG027

<table>
<thead>
<tr>
<th>Email Address</th>
<th>2020</th>
<th>2019</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>arabic.arabic@arabic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="mailto:chinese@chinese.chinese">chinese@chinese.chinese</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="mailto:chinese@ascii.ascii">chinese@ascii.ascii</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="mailto:ascii@chinese.ascii">ascii@chinese.ascii</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="mailto:ascii@chinese.newlong">ascii@chinese.newlong</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="mailto:ascii@ascii.newshort">ascii@ascii.newshort</a></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Estimated Support of EAI in Email Systems Under All TLDs

Only 9.7% of the domains sampled were EAI ready in 2019.

This is based on mail servers found through MX records in zones of all top-level domains (TLDs).

For details on methodology, see UASG021D: EAI Readiness in TLDs
Technology Stack for UA Consideration

Applications and Websites
- Wikipedia.org, ICANN.org, Amazon.com, custom websites globally
- PowerPoint, Google-Docs, Safari, Acrobat, custom apps

Social Media and Search Engines
- Chrome, Bing, Safari, Firefox, local (e.g., Chinese) browsers
- Facebook, Instagram, Twitter, Skype, WeChat, WhatsApp, Viber

Programming Languages and Frameworks
- JavaScript, Java, Swift, C#, PHP, Python
- Angular, Spring, .NET core, J2EE, WordPress, SAP, Oracle

Platforms, Operating Systems and System Tools
- iOS, Windows, Linux, Android, App Stores
- Active Directory, OpenLDAP, OpenSSL, Ping, Telnet

Standards and Best Practices
- IETF RFCs, W3C HTML, Unicode CLDR, WHATWG
- Industry-based standards (health, aviation, ...)

UA Readiness needs to be checked and fixed (as needed) for multiple frameworks, utilities, tools, and applications at multiple layers of technology
All email agents must be configured to send and receive internationalized email addresses. See EAI: A Technical Overview for details.

- **MUA** – Mail User Agent: A client program that a person uses to send, receive, and manage mail.

- **MSA** – Mail Submission Agent: A server program that receives mail from a MUA and prepares it for transmission and delivery.

- **MTA** – Mail Transmission Agent: A server program that sends and receives mail to and from other Internet hosts. An MTA may receive mail from an MSA and/or deliver mail to an MDA.

- **MDA** – Mail Delivery Agent: A server program that handles incoming mail and typically stores it in a mailbox or folder.
Quiz
To enhance systems to be Universal Acceptance (UA) ready, which of the following categories of domain names and email addresses are relevant?

1. ASCII domain names.
2. Internationalized Domain Names (IDNs).
3. Internationalized email addresses (EAI).
4. All the above.
5. Only 2 and 3.
Fundamentals for Internationalized Domain Names and Email Addresses
Unicode Encoding

- Unicode encodes glyphs into codepoints for different scripts of the world.
  - Codepoints shown in hex using the U+XXXX notation.
  - Unicode files typically in UTF8 format, using a variable number of bytes for a codepoint.
  - ASCII is used as is in Unicode: \(e = \text{ASCII 65} = \text{U+0065}\).

- There are multiple ways to encode certain glyphs in Unicode:
  - \(\text{è} = \text{U+00E8}\)
  - \(e + ` = \text{è} = \text{U+0065 + U+0300}\)

- Normalization ensures that the end representation is the same, even if users type differently.
  - IDN standards recommend using Normalization Form C (NFC).
  - Generates \text{U+00E8} for both input versions above.
A domain name is an ordered set of labels or strings: www.example.co.uk.
- The top-level domain (TLD) is the rightmost label: "uk"
- Initially, TLDs were only two or three characters long (e.g., .ca, .com).
- Now TLDs can be longer strings (e.g., .info, .google, .engineering).
- TLDs delegated in the root zone can change over time, so a fixed list can get outdated.

Domain names can also be internationalized when one of the labels contains at least one non-ASCII character.
- For example: www.example.ca or 普遍接受-测试.世界.

Use the latest IDN standard called IDNA2008 for IDNs.
- Do not use libraries for the outdated IDNA2003 version.
There are two equivalent forms of IDN domain labels: U-label and A-label.

- Human users use the IDN version called U-label (using UTF-8 format): exâmple
- Applications or systems internally use an ASCII equivalent called A-label:
  1. Take user input and normalize and check against IDNA2008 to form IDN U-label.
  2. Convert U-label to punycode (using RFC3492).
  3. Add the “xn--” prefix is added to identify the ASCII string as an IDN A-label.
    - exâmple => exemple-xta => xn--exmple-xta.
    - 普遍接受-测试 => --f38am99bqvcd5liy1cxsg => xn----f38am99bqvcd5liy1cxsg.

Email address syntax: mailboxName@domainName

- EAI has the mailboxName in Unicode (in UTF8 format).
- The domainName can be ASCII or IDN.
  - For example: kévin@example.org or すし@快手.游戏.
Universal Acceptance Issues

- Some applications are still verifying domain names incorrectly by using one of the outdated methods:
  - Check for a fixed length of TLD between 2-4 characters (TLD can be up to 63 characters).
  - Check from a fixed set of TLDs, e.g., using static list of strings.
  - Check for only ASCII characters.

- Some applications do not cater to additional requirements for validating EAI:
  - Check mailbox name to be a valid string in UTF-8 format.
  - DomainName can be ASCII or IDN.
UA Readiness Testing Framework

- Based on UASG026, the application components can be generalized to put emphasis on the processing of internationalized identifiers.

- Each gate has its own set of requirements and processing.

  - AT: Accept test
  - VT: Validate test
  - P1T: Process test on the input
  - ST: Store test
  - P2T: Process test on the output
  - DT: Display test
Validating User Input

Validating user input, or any input, is extremely useful for various reasons, some of which include: a better user experience, increased security, and avoiding irrelevant issues.

Validating domain names and email addresses is useful.

Some validation methods for domain names and email addresses:
- Basic syntax checks: is the syntax of the string correct?
  - Does the domain name contain ‘.’?
  - Does the email address contain ‘@’ and a valid domain name part?
- Functional checks: does the domain name or email address work?
  - Is the top-level domain (TLD) in use?
  - Is the whole domain name in use?
  - Is the email in use?
Validating Domain Name

- Validating syntax:
  - ASCII: RFC1035
    - Composed of letters, digits, and hyphen.
    - Max length is 255 octets with each label up to 63 octets.
  - IDN: IDNA2008 (RFCs 5890-5894)
    - Valid A-labels
    - Valid U-labels

- Validating function:
  - Is the top-level domain (TLD) in use?
    - Verify against the list of TLDs.
    - Verify using a DNS request.
  - Is the whole domain name in use?
    - Verify using a DNS request.
Resolving Domain Name

- After validation, a software would then use the domain name identifier as:
  - A domain name to be resolved in the DNS.

- Therefore, to be UA compliant, the software has to use proper methods that support UA.
  - For example, passing a U-Label to the traditional functions or methods may not succeed, as it is not expecting a UTF8 domain name.
Validating Email Address

- An email address is composed of: mailboxName@domainName

- Validating syntax:
  - For domainName, see earlier discussion.
  - For mailboxName:
    - ASCII
    - UTF8 (for EAI)

- Validating function:
  - Is the domain name set up to send and receive emails?
  - Is the mailbox name able to send and receive emails?
Sending Email

- After validation, a software would then use the email identifier as:
  - An email-address based user id.
  - To send an email.

- Therefore, to be UA compliant, the software must use proper methods that support UA.
  - For example, passing an UTF8 mailbox name email address to a mail sender may not succeed, as it is not expecting a UTF8 mailbox name in the email address.
Email Regular Expressions (Regex)

- **Basic: something@something**
  - `^(.+@.+)$`

- From [owasp.org](http://owasp.org) (security):
  - `^[^a-zA-Z0-9_+&*-]+(?:\:[^a-zA-Z0-9_+&*-]+)*@[^\x00-\x08^\x0b-\x0c^\x0e-\x1f]+\.[a-zA-Z]{2,7}$`
    - Does not support EAI, i.e., mailbox name in UTF-8 not allowed: `[a-zA-Z0-9_+&*-]`
    - Does not support ASCII TLD longer than 7 characters: `[a-zA-Z]{2,7}`
    - Does not support U-labels in IDN TLD: `[a-zA-Z]`
  - But OWASP is THE reference for security.
    - Therefore, you may end up fighting with your security team to use a UA-compatible Regex instead of the “standard” one from OWASP.
Sending Email

- A comprehensive list of UA test cases is documented in [UASG004](#).
- Developers are strongly encouraged to use these test cases in its unit and system testing.
A company built a website where international consumers can subscribe via their email. Since the subscription form is user input, developers validated the email address before trying to send the email.

- Developers went to StackOverflow and found a regular expression (regex) to perform the validation:

```java
public static final Pattern VALID_EMAIL_ADDRESS_REGEXP =
Pattern.compile("^[A-Z0-9._%+-]+@[A-Z0-9.-]+\.[A-Z]{2,6}$", Pattern.CASE_INSENSITIVE);

public static boolean validate(String emailStr) {
    Matcher matcher = VALID_EMAIL_ADDRESS_REGEXP.matcher(emailStr);
    return matcher.find();
}
```

- The regex limits mailbox to letters A-Z, digits 0-9, and some symbols, the domain labels to letters, digits and hyphen, and the top-level domain to letters only with length 2-6.

Would this regex work for the company’s website? Why or why not?
<table>
<thead>
<tr>
<th>LANGUAGE</th>
<th>LIB NAME</th>
<th>COMPLIANCE (%)</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Javascript</td>
<td>Idna-Uts46</td>
<td>85.5</td>
<td>IDN</td>
</tr>
<tr>
<td>Javascript</td>
<td>Nodemailer</td>
<td>84.3</td>
<td>Mail</td>
</tr>
<tr>
<td>Javascript</td>
<td>Validator</td>
<td>94.2</td>
<td>Mail</td>
</tr>
<tr>
<td>Python3</td>
<td>Django Auth</td>
<td>48.1</td>
<td>Mail</td>
</tr>
<tr>
<td>Python3</td>
<td>Email Validator</td>
<td>86.3</td>
<td>Mail</td>
</tr>
<tr>
<td>Python3</td>
<td>Encodings_Idna</td>
<td>67.7</td>
<td>IDN</td>
</tr>
<tr>
<td>Python3</td>
<td>Idna</td>
<td>100</td>
<td>IDN</td>
</tr>
<tr>
<td>Python3</td>
<td>Smtplib</td>
<td>84.3</td>
<td>Mail</td>
</tr>
<tr>
<td>Rust</td>
<td>Idna</td>
<td>87.1</td>
<td>IDN</td>
</tr>
<tr>
<td>Rust</td>
<td>Lettre</td>
<td>7.8</td>
<td>Mail</td>
</tr>
<tr>
<td>C</td>
<td>Libcurl</td>
<td>84.3</td>
<td>Mail</td>
</tr>
<tr>
<td>C</td>
<td>Libidn2</td>
<td>95.2</td>
<td>IDN</td>
</tr>
<tr>
<td>C#</td>
<td>Mailkit</td>
<td>84.3</td>
<td>Mail</td>
</tr>
<tr>
<td>C#</td>
<td>Microsoft</td>
<td>83.9</td>
<td>IDN</td>
</tr>
<tr>
<td>Go</td>
<td>Mail</td>
<td>100</td>
<td>Mail</td>
</tr>
<tr>
<td>Go</td>
<td>Idna</td>
<td>79</td>
<td>IDN</td>
</tr>
<tr>
<td>Go</td>
<td>Smtpl</td>
<td>19.6</td>
<td>Mail</td>
</tr>
<tr>
<td>Java</td>
<td>Commons-Validator</td>
<td>85.5</td>
<td>Mail, IDN</td>
</tr>
<tr>
<td>Java</td>
<td>Guava</td>
<td>77.8</td>
<td>IDN</td>
</tr>
<tr>
<td>Java</td>
<td>ICU</td>
<td>93.5</td>
<td>IDN</td>
</tr>
<tr>
<td>Java</td>
<td>Jakartamail</td>
<td>82.4</td>
<td>Mail</td>
</tr>
<tr>
<td>Java</td>
<td>JRE</td>
<td>71</td>
<td>IDN</td>
</tr>
</tbody>
</table>
○ Be aware that UA identifiers may not be fully supported in software and libraries.
○ Use the right libraries and frameworks.
○ Adapt your code to properly support UA.
○ Do unit and system testing using UA test cases to ensure that your software is UA ready.
Email Address Internationalization (EAI)
EAI Protocol Changes

- **SMTP**
  - Is augmented to support EAI.
  - Has a signaling flag (SMTPUTF8) to specify support of EAI.
  - All SMTP servers in the path must support EAI to successfully deliver the email.

- **POP/IMAP**
  - Are augmented to properly support EAI.
  - Have a signaling flag to specify support of EAI.
SMTPUTF8 Example

Server S forwarding an email to server R

S: <connect>
R: 220 receive.net ESMTP
S: EHLO sender.org
R: 250-8BITMIME
R: 250-SMTPUTF8
R: 250 PIPELINING
S: MAIL FROM:<猫王@普遍接受-测试.世界> SMTPUTF8
R: 250 Sender accepted
S: RCPT TO:<ray@receive.net>
R: 250 Recipient accepted

Specific SMTPUTF8 Signaling (EAI support)
SMTPUTF8 Example

S: DATA
R: 354 Send your message
S: From: 猫王 <猫王@普遍接受-测试.世界>
S: To: ray@receive.net
S: Subject: 我们要吃午饭吗？
S:
S: How about lunch at 12:30?
S:
R: 250 Message accepted 389dck343fg34
S: QUIT
R: 221 Sayonara
Protocol Changes, Delivery Path Considerations

To send and receive an email with EAI:
- All email parties involved in the delivery path have to be updated for EAI support.
- If a single SMTP server in the path does not support EAI, then the email is not delivered.
Protocol Changes, Delivery Path Considerations

- What happens when one email (SMTP) server in the path does not support EAI?
  - The last server trying to send to the next hop:
    - Sends back to the sender user a report of unable to deliver.
    - Drops the email.
  - Similar to reports that a sender receives when an email address does not exist.
Considerations

- Case folding:
  - In ASCII, email users expect the equivalence of lowercase and uppercase.
    - Example: PETER@example.com and peter@example.com will be delivered to the same mailbox.
  - Typically for EAI, the case folding functionality is not implemented in most EAI-ready software.

- SPAM:
  - EAI emails may be considered as spam by spam filtering software even when proper records are enabled.

- Software/Services:
  - Not every server/client software and services support EAI.
Components of Email System:
- Mail User Agent
- Mail Submission Agent
- Mail Transfer Agent
- Mail Delivery Agent
- Mail Service Provider

**L1** - EAI level 1 - sends to and receives from EAI addresses
**L2** - EAI level 2 - L1 plus provides local EAI addresses

<table>
<thead>
<tr>
<th>Name</th>
<th>MUA</th>
<th>MSA</th>
<th>MTA</th>
<th>MDA</th>
<th>MSP</th>
<th>Web mail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coremail</td>
<td>Few</td>
<td>All L2</td>
<td>Most L2</td>
<td>Few</td>
<td>All L2</td>
<td>Most L2</td>
</tr>
<tr>
<td>MS Outlook.com</td>
<td>Most L1</td>
<td>Most L1</td>
<td>Most L1</td>
<td>None</td>
<td>None</td>
<td>Most L1</td>
</tr>
<tr>
<td>Yandex Mail</td>
<td>Few</td>
<td>None</td>
<td>None</td>
<td>Few</td>
<td>Part</td>
<td>Few</td>
</tr>
<tr>
<td>Roundcube</td>
<td>Most L2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apple Mail</td>
<td>Few</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mozilla</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thunderbird</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS Outlook</td>
<td>Most L1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS Exchange Server (hosted)</td>
<td>All L1</td>
<td>All L1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exim</td>
<td>Most L2</td>
<td>All L2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postfix</td>
<td>All L2</td>
<td>All L2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sendmail</td>
<td>Not tested</td>
<td>Not tested</td>
<td>Not tested</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fetchmail</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Not tested</td>
</tr>
<tr>
<td>Courier</td>
<td>All L2</td>
<td>All L2</td>
<td>All L2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gmail</td>
<td>All L1</td>
<td>All L1</td>
<td>All L1</td>
<td>Few</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XgenPlus</td>
<td>Not tested</td>
<td>Not tested</td>
<td>Not tested</td>
<td></td>
<td>All L2</td>
<td>Not tested</td>
</tr>
</tbody>
</table>
Are Your Software Applications UA Ready?
ICANN’s Journey to UA Readiness - Model

- Stage 1: Update services to support both new short and long ASCII TLDs.
- Stage 2: Update services to support non-ASCII Internationalized Domain Names (IDNs) in Unicode (U-label), and ASCII-based IDN representations in Punycode (A-label).
- Stage 3: Update infrastructure and services to support non-ASCII email addresses.
  - Note: all components must support Email Address Internationalization (EAI) before infrastructure is compliant.

See details in ICANN’s Case Study
Next Steps and Community Support

- UASG and ICANN continue to undertake gap analysis, remediation, training and outreach:
  - Gap analysis – Social Media, Browsers, Programming Languages, EAI Tools, etc.
  - Remediation – engaging technology forums (e.g. Github) and bug reporting.
  - Training and outreach – through local initiatives and ambassadors.

We request the community to help address UA readiness and lead by example:

1. **Raise awareness** of the technical problems within the community.

2. **Upgrade and use UA ready systems** as a community to create the necessary demand, e.g. upgrade email servers, use email in local language.

3. **Advocate more broadly** to support UA in their systems (e.g. in e-govt. services; the private sector organizations, etc.).

Such activities may be undertaken in collaboration with UA Local Initiative and UA Ambassadors.
Some Relevant Materials

- See [https://uasg.tech](https://uasg.tech) for a complete list of reports.
  - Universal Acceptance Quick Guide: [UASG005](#)
  - Introduction to Universal Acceptance: [UASG007](#)
  - Quick Guide to EAI: [UASG014](#)
  - EAI – A Technical Overview: [UASG012](#)
  - EAI – Evaluation of Major Email Software and Services: [UASG021B](#)
  - Universal Acceptance Readiness Framework: [UASG026](#)
  - Considerations for Naming Internationalized Email Mailboxes: [UASG028](#)
  - UA Readiness Report 2020: [UASG029](#)
  - Evaluation of EAI Support in Email Software and Services Report: [UASG030](#)
  - Frequently Asked Questions (FAQs): UA Readiness of Programming Languages and Email Tools: [UASG031](#)

- Please email [info@uasg.tech](mailto:info@uasg.tech) or [UAProgram@icann.org](mailto:UAProgram@icann.org) for further information.
Get Involved!
Get Involved!

- Join APAC EAI Implementers’ Group mailing list for technical support (by THNIC)

- For more information on UA, email info@uasg.tech or UAProgram@icann.org

- Access all UASG documents and presentations at: https://uasg.tech

- Access details of ongoing work from wiki pages: https://community.icann.org/display/TUA

- Register to participate or listen in the UA discussion list at: https://uasg.tech/subscribe

- Register to participate in UA working groups here.