INTERNATIONALIZED DOMAIN NAMES
-A BASIC INTRODUCTION
2 November 2008

Tina Dam, Director, IDN Program
tina.dam@icann.org

باهر عصمت، مدير علاقات إقليمية - الشرق الأوسط
baher.esmat@icann.org
Agenda

- IDN definitions and basics
- How does IDNs work
  - including examples of applications and their various implementations
- IDN wiki facility
- IDN TLD introduction processes
- IDN TLD technical string requirements
Characters in the DNS

- Search on "US-ASCII character set"
- The DNS can handle all US-ASCII characters
  - Examples:
    - (a…z), (0…9), (-)
    - ( ) SPACE
    - ( !) EXCLAMATION MARK
    - ("") QUOTATION MARK
    - (#) NUMBER SIGN
    - ($) DOLLAR SIGN
    - (%) PERCENT SIGN
    - (&) AMPERSAND
    - (') APOSTROPHE
Characters, DNS, and domain names...

- All TLD registries have implemented the LDH rule
  - Domain names can only contain:
    - (a,b,...z)
    - (0,1,...9)
    - (-)
- That was before internationalization....
Internationalized domain names are:
- Names with characters other than the standard ASCII (a,b,...z), (0,1,...9), (-)

Example:
\[ xn--ngbrx4e.tld \leftrightarrow tld.عربية \]

IDNs are about localized solutions
- But need to be ‘internationalized’ due to the global nature of the Internet
IDNs - What we have / what we need

- IDNs have existed as second level since 2003
  - under web protocol standards
  - email protocol standards are underway (IETF)
- We also need IDN TLDs
  - 北京.中国; [xn--1lq9oi.xn--fiQs8S]

<table>
<thead>
<tr>
<th>Domain Availability Today</th>
<th>ASCII domain names (a, b, ..., z), (0,1, ...,9), (-)</th>
<th>domainname.TLD icann.org</th>
</tr>
</thead>
<tbody>
<tr>
<td>Future Addition</td>
<td>IDN second level</td>
<td>实例.TLD under various existing TLDs</td>
</tr>
<tr>
<td></td>
<td>IDN TLDs</td>
<td>실례. 테스트</td>
</tr>
</tbody>
</table>
Why Internationalization?

- DNS handling US-ASCII character set
  - a natural choice at the time
  - no expectation to current commercial value
  - Unicode was not available

- IDNs a natural expansion for global usability
  - allow users to use domain names in local scripts
  - no need to learn US-ASCII, user choice
    - some applications have implemented IDNA
    - still need internationalization of TLD
IDNA – Protocol Functionality

• Domain Name Resolution Process:

IDNA is a client based protocol:
1. User types in 실례.test in for example a browser
2. 실례.test gets converted to Unicode, if not already entered as such
3. IDNA conversion \(\rightarrow\) xn--9n2bp8q.test
Displayed Form vs. Stored Form

- Historically the domain name you register is also the domain names stored and usable in the DNS

- This is changed with introduction of IDNs

- The stored form usually gives no meaning
  - Example: فرسالنهر.tld → xn--mgbtbg2evoai.tld

- However, there are exceptions:
  - xn--gibberish - decodes into the Arabic characters ﺖ٩٧٨٩٠١٢٣٤٥٦٧٨٩
  - xn--trademark - with different versions of trademarks
  - This is coincidentally and hence not intentionally, but implementations showing xn– to the user is a concern

- xn-- prefix indicates to application software that the label needs to be decoded back into Unicode for proper display to the user
IDNA protocol and IDN wiki – try it out

- http://idn.icann.org
- http://josefsson.org/idn.php
  - IDNA ToASCII
  - IDNA ToUnicode

- If you can’t type in an IDN then search for your favorite newspaper online and copy-paste it
  - or you can copy-paste text from the IDN wiki

- Try copy / paste between applications you normally use
<table>
<thead>
<tr>
<th>Script</th>
<th>Language</th>
<th>SLD.TLD U-labels</th>
<th>SLD A-label</th>
<th>TLD A-label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic</td>
<td>Arabic</td>
<td>xn--mgbh0fb</td>
<td>xn--kgbechtv</td>
<td></td>
</tr>
<tr>
<td>Arabic</td>
<td>Persian</td>
<td>xn--mgbh0fb</td>
<td>xn--hgbkk6ag7j3bba</td>
<td></td>
</tr>
<tr>
<td>Chinese, simplified</td>
<td>Chinese</td>
<td>xn--f8q00a</td>
<td>xn--4zwm56d</td>
<td></td>
</tr>
<tr>
<td>Chinese, traditional</td>
<td>Chinese</td>
<td>xn--f8q00a</td>
<td>xn--g6w251d</td>
<td></td>
</tr>
<tr>
<td>Cyrillic</td>
<td>Russian</td>
<td>xn--e1afmikfd</td>
<td>xn--80akhbylnaj4f</td>
<td></td>
</tr>
<tr>
<td>Devanagari</td>
<td>Hindi</td>
<td>xn--p1b6c1b4b3a</td>
<td>xn--11b5bx3a0aj6g</td>
<td></td>
</tr>
<tr>
<td>Greek</td>
<td>Greek</td>
<td>xn--hxaajhls2aaz3al</td>
<td>xn--jsaxpdlp</td>
<td></td>
</tr>
<tr>
<td>Hangul</td>
<td>Korean</td>
<td>xn--9nq2p8q</td>
<td>xn--9q4b11y5a</td>
<td></td>
</tr>
<tr>
<td>Hebrew</td>
<td>Yiddish</td>
<td>xn--rdbk5d8ap9r8a8d</td>
<td>xn--deba0ad</td>
<td></td>
</tr>
<tr>
<td>Kamit Hirigana, and Katakana</td>
<td>Japanese</td>
<td>xn--r8z44g</td>
<td>xn--zckzah</td>
<td></td>
</tr>
<tr>
<td>Tamil</td>
<td>Tamil</td>
<td>xn--zko6cc5bi7f6e</td>
<td>xn--bhef6aya9esc7a</td>
<td></td>
</tr>
</tbody>
</table>
Status of the .test wiki

- **Purpose of the IDNwiki:**
  - Introduce users to IDN TLDs
  - Applications test environment for usability
  - Registry information about user problems
- **Conduct an experiment with IDN TLDs**
  - not a pre-requisite for production in root zone
  - no registrations are available
- **Functions as a “normal wiki”, user access**
## IDN TLD Introduction Processes

### Implementation of Country-code IDN TLDs – Fast Track
- Implementation of recommended policy from IDNC WG
- Review of contractual elements and relationship to IDN TLD operators
- Review of technical requirements for IDN TLD management

### Country-code IDN TLDs – Long Term
- Full policy that caters for all
- Follows the full ccNSO Policy Development Process

### New Generic TLDs
- New ongoing policy for new gTLDs
- Includes internationalized domains
- Focus on non-ASCII squatting & confusingly similarity solutions
IDN TLD Technical Requirements

- The label must be a valid internationalized domain name, as specified in technical standards http://www.icann.org/en/topics/idn/rfcs.htm. This includes the following, non-exhaustive, list of limitations:
  - Must only contain Unicode code points that are defined as “Valid” in The Unicode Codepoints and IDNA (Internet Draft “draft-faltstrom-idnabis-tables”)
  - Must be fully Normalization Form C compliant, as described in Unicode Standard Annex #15: Unicode Normalization Forms.
  - Must not contain characters with a mixture of directionality properties.
  - Must not contain any leading or trailing digits (of any kind of digits).
IDN TLD Technical Requirements II

- The label must meet the relevant criteria of the ICANN Guidelines for the Implementation of Internationalized Domain Names. This includes the following, non-exhaustive, list of limitations:
  
  - All code points in a single label will be taken from the same script as determined by the Unicode Standard Annex #24: Unicode Script Property.
  
  - Exceptions are permissible for languages with established orthographies and conventions that require the commingled use of multiple scripts. However, even in the case of this exception, visually confusible characters from different scripts will not be allowed to co-exist in a single set of permissible code points unless a corresponding policy and character table is clearly defined.
Internationalization of the internet means that the internet is equally accessible from all languages and scripts.
شكرًا لكم

tina.dam@icann.org
http://icann.org/topics/idn