

# International Domain Names in IE7

Michel Suignard  
Microsoft Corp

# IDN implementation basics

- Major part in network stack (not application)
- Minimize Punycode exposure in the User Interface
- Equivalency between Punycode and Unicode links (IRI)
- Compatibility (legacy support)
- Addressing security concerns

# Conversion to/from punycode

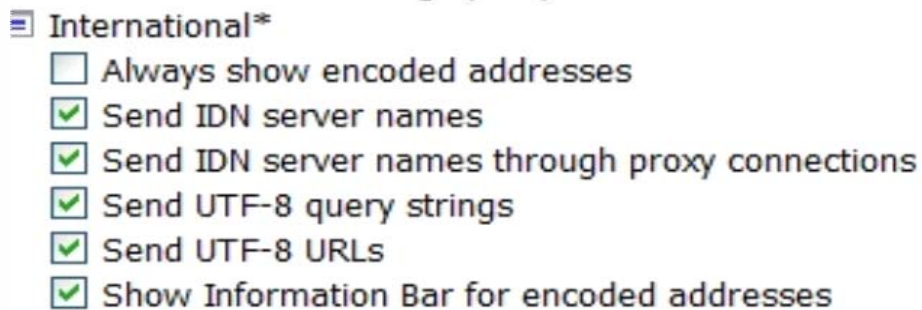
- New Windows functions:
  - IdnToASCII
  - IdnToUnicode
  - IdnToNameprepUnicode (similar to IdnToASCII with the final Punycode step discarded)
- Available in Vista and bundled with IE7
- Also available in MS download site along with script based API

# Link consistency

- A hostname entered as punycode is treated as it was entered in Unicode
  - räksmörgås.josefsson.org is the same as xn--rksmrgs-5wao1o.josefsson.org
  - 納豆.w3.mag.keio.ac.jp is the same as xn--99zt52a.w3.mag.keio.ac.jp
- Names are entirely processed in native Unicode encoding until they are sent to the wire
  - But may be shown in Punycode format if deemed 'suspicious'

# Compatibility

- Detailed controls in the Advanced options dialog allowing to get previous behavior (still experimental)



# Preventing homograph spoofing attacks

- Scripts unknown to a user are presumed suspicious
- Mixed of scripts within a label are suspicious, unless customary for a given language/locale
- Suspicious sites are presented in Punycode form, not Unicode in all UI elements

# Script based API

- Script detection:
  - [Downlevel]GetStringScripts returns script list for a string
- Locale based script list
  - GetLocaleInfo(..., LC\_SCRIPTS,...) returns script list for a locale (Vista)
  - DownlevelGetLocaleScripts (down level)
- Script comparisons
  - [Downlevel]VerifyScripts compares two lists of scripts

# More work to do

- In-script confusables
  - Opt in servicing model
  - In script confusable names can be compared to a list of known non-phishing hostnames.
- UI model may use color coding to show suspicious levels (green, yellow, red) along with explicit markers.
- Usage of the restricted list from UTS 39:
  - Idnchars.txt
  - Virtually no symbols
  - Only modern use characters



# Resources

- [http://msdn.microsoft.com/library/en-us/intl/nls\\_IDN.asp](http://msdn.microsoft.com/library/en-us/intl/nls_IDN.asp)
  - IdnDL.dll (down level scripts APIs)
  - Normaliz.dll (IDN and normalization functions)
- <http://blogs.msdn.com/ie/>
  - Recent developments on IE7
- <http://www.unicode.org/reports/tr36/tr36-4.html>
  - Information about confusables, IDNchars.txt, ...
- <http://www.unicode.org/reports/tr24/tr24-7.html>
  - Script tables

**DEMO!**