

IDNAbis WG Status

April 2009

John C Klensin

ASIWG, 2009.04.02

When will we be finished?

- Optimistic view:
 Within the next month or so
- Pessimistic view:
 I believed that in Dubai, too.

What is the problem?

- Two issues of principle
 - IDNA2003 compatibility and transition
 - Lookup-time checks
- Many per-character issues
 - WG has trouble deciding when
 - Script community appears to disagree
 - WG cannot figure out who is an expert
 - Different people claim expertise but do not agree
 - Results when WG cannot decide are unpredictable but may tend to follow Unicode recommendations.

IDNA2003 Compatibility

- Some people in WG that believe either that
 - No change from IDNA2003 is needed at all
 - No change should be made other than to update IDNA2003 tables for characters that are new after Unicode 3.2
 - Strict backward-compatibility is a requirement
 - If it works in IDNA2003, even in non-conforming implementations, it must work forever.
- *WG seems* to have decided against these positions

IDNA2003 Transition

- What to do about IDNA2003 registrations and lookups?
 - Process all of them on lookup, including disallowed characters?
 - Process some of them?
 - Preserve all mappings?
 - Quick, painful, transition?
- Is this “transition” or permanent retention of IDNA2003?

Lookup-time Checks

- Some people in WG believe that there should be few, if any, checks at lookup time.
- Extreme version of “leave it all to the registries”

Tatweel as an Example

- The Tatweel situation illustrates several of these issues (it is *not* the only one)
- Specifically...
 - It appears that the script community does not agree about what should be done
 - Different people who claim to be experts are taking different positions
 - The vast majority of the active participants in the WG have no real experience with the script or languages written in it
 - IDNA2003 allows the character

Digits are even more complicated

- WG sees apparently-inconsistent explanations of what is needed.
- Unicode explanation of separation of Arabic-Indic digit types is hard to understand
- Use of ۝ (U+06F2) as a replication indicator in Jawa with European digits in numerals complicates a rule against Arabic-Indic mixing with European digits.

When will this end?

I really, really, wish I knew.