

Subject: UTC Agenda Item: Script_Extensions: Short name, and regex range semantics
From: Markus Scherer <markus.icu@gmail.com>
Date: Mon, 26 Jul 2010 13:03:50 -0700
To: l2doc@unicode.org, utc-chair@unicode.org

Please add the following to the document registry and to the UTC agenda.

Thanks,
 markus

 Script_Extensions: Short name, and regex range semantics

Background: Unicode 6.0 adds a provisional property Script_Extensions with a new data file, ScriptExtensions.txt. (Documented in the UAX #24 update.)

Request:

1. I request that the UTC document a short name (property alias) for Script_Extensions.
 - 1a. I suggest "scx" as the short name.
2. Please also document that when this property is implemented in character ranges (e.g., [:Script_Extensions=Arab:] in regular expressions or UnicodeSet), the range should include characters that (a) have the specified Script property value, *or* (b) where the specified Script property value occurs in the Script_Extensions. (inclusive-or)

Rationale:

Unicode properties are implemented in regular expressions, UnicodeSet, UCD XML files and other text data formats, and the short names are often used for convenience. Documenting the short name prevents different implementations from inventing different short names which would lead to unnecessary incompatibilities.

Similarly, non-obvious semantics should be documented for consistent implementations. The proposed semantics anticipate "is commonly used in script x" as typical usage. If the simple regex range did not include the regular Script value, then users would always have to remember to use the union of two regex ranges like [[:sc=Arab:][:scx=Arab:]].

Provisional property:

Given that the Script_Extensions data is provisional, the property aliases should be documented as such. For example in PropertyAliases.txt:

```
# =====
# Miscellaneous Properties
# =====
scx ; Script_Extensions # provisional: may be changed or removed in a future version
```

Or else in ScriptExtensions.txt:

```
# Provisional short and long property names:
# scx ; Script_Extensions
```