

Update Oniguruma for Unicode 6

10/27/2010 09:20 AM - runpaint (Run Paint Run Run)

Status:	Closed	
Priority:	Normal	
Assignee:		
Target version:		
Description =begin Unicode 6.0 has been released, so it would be preferable to update Oniguruma for Ruby 1.9.3. A patch is attached which updates enc/unicode/unicode/name2ctype.kwd, and fixes the instructions in tool/enc-unicode.rb. Given the scope of this update, I'd prefer this merged sooner rather than later so we have time to test it. A tentative test follows: U+20B9, INDIAN RUPEE SIGN, is a new codepoint in 6.0. It has the general category Sc. <code>^p{sc}/u =~ "\u{20B9}" #=> 0</code> U+0B72, ORIYA FRACTION ONE QUARTER, is a new codepoint in the Oriya script. <code>^p{oriya}/u =~ "\u{0b72}" #=> 0</code> U+0B77, ORIYA FRACTION THREE SIXTEENTHS, is the last codepoint in the Oriya script block. <code>^p{oriya}/u =~ "\u{0b78}" #=> nil</code> U+FBC0, TAMIL VOWEL SIGN II, is a new codepoint in the Arabic Presentation Forms-A block. <code>^p{arabic}/u =~ "\u{fbc0}" #=> 0</code> U+1F130, SQUARED LATIN CAPITAL LETTER A, is a new codepoint with a general category of So. <code>^p{so}/u =~ "\u{1f130}" #=> 0</code> U+0847, MANDAIC LETTER IT, is a new codepoint in the Mandaic script. The Mandaic script is new. <code>^p{mandaic}/u =~ "\u{0847}" #=> 0</code>		

(When this is complete, NEWS needs to be updated, too).
=end

Associated revisions

Revision **fc9176ac0e0edd4ba35059768cb2c1bdd3ec529b** - 10/28/2010 11:14 AM - naruse (Yui NARUSE)

- tool/enc-unicode.rb,
enc/unicode/name2ctype.h, enc/unicode/name2ctype.h.blk,
enc/unicode/name2ctype.kwd, enc/unicode/name2ctype.src:
Update Oniguruma for Unicode 6.
patched by Run Paint Run Run. [ruby-core:32923] #3989

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@29620 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

History

#1 - 10/28/2010 09:49 AM - runpaint (Run Paint Run Run)

=begin
I'm working on a test suite for this. So far, I've found some scripts that don't appear to be recognised (may also exist in trunk; I haven't checked yet):

1. (0xa840..0xa877).select{|o| o.chr('utf-8') =~ /\p{Phag}/} #=> [] http://en.wikipedia.org/wiki/Phags-pa_script
2. (0x1c00..0x1c23).select{|o| o.chr('utf-8') =~ /\p{Lepcha}/} #=> [] http://en.wikipedia.org/wiki/Lepcha_script
3. (0x103a0..0x103c3).select{|o| o.chr('utf-8') =~ /\p{xpeo}/} #=> [] http://en.wikipedia.org/wiki/Old_Persian_cuneiform
4. The "Unknown" script, defined in http://unicode.org/reports/tr24/#Special_Explicit, isn't handled by enc-unicode.rb. /\p{Zzzz}/, for example, raises a SyntaxError.

=end

#2 - 10/28/2010 08:17 PM - naruse (Yui NARUSE)

- Status changed from Open to Closed
- % Done changed from 0 to 100

=begin
This issue was solved with changeset r29620.
Run Paint, thank you for reporting this issue.
Your contribution to Ruby is greatly appreciated.
May Ruby be with you.

=end

Files

onig-u6.patch	731 KB	10/27/2010	runpaint (Run Paint Run Run)
---------------	--------	------------	------------------------------