

Ruby - Bug #7442

StringScanner#charpos vs StringScanner#pos

11/27/2012 11:04 AM - zenspider (Ryan Davis)

Status:	Closed	Backport:
Priority:	Normal	
Assignee:		
Target version:	3.0	
ruby -v:	1.9.x	
Description		
<p>=begin</p> <p>I talked to Matz at rubyconf and he agreed this was a bug I should file. Sorry I took so long to do so.</p> <p>As mentioned in #3482, StringScanner#pos is byte-oriented even when scanning multibyte strings. The reasoning was that IO#pos is byte-oriented so this is to spec and functioning correctly. The problem is that StringScanner isn't <i>just</i> an IO as it also represents a String and the progress scanning through it. Strings in 1.9+ must respect their encodings and with a few exceptions don't even support the idea of naked bytes. I think StringScanner must be able to respect that.</p> <p>Given that ss is a StringScanner instance on a string with a valid encoding, getting the substring of the current progress via ss.string[0..ss.pos] can result in a String with <i>invalid</i> encoding. I propose that we add #charpos to make it possible to pull out a valid substring. This would also be useful towards being able to report proper offset or column information in the case of an error when you're using StringScanner as your lexer.</p> <p>This is the code that I needed to get proper char-offsets (and substrings--I needed both for my purposes):</p> <pre>def string_to_pos string.byteslice(0, pos) end def charpos string_to_pos.length end =end</pre>		
Related issues:		
Related to Ruby - Feature #1159: StringScanner <code>StringScanner#charpos</code>		Rejected 02/14/2009

History

#1 - 11/27/2012 12:25 PM - mame (Yusuke Endoh)

- Status changed from Open to Feedback
- Target version set to 3.0

Sorry, it is too late to fix such a spec-level bug. Setting the target to Next Major.
If you create and commit a patch by preview2 (1 Dec.), and if it does not lead to any problem (and any discussion) at all, we might include it in 2.0.0.

--
Yusuke Endoh mame@tsg.ne.jp

#2 - 11/28/2012 09:31 AM - zenspider (Ryan Davis)

Committed revision 37916.

Please beat up on it.

#3 - 11/29/2012 09:23 AM - zenspider (Ryan Davis)

No objections (yet)... can this be merged to 2.0 branch for next preview release?

#4 - 11/29/2012 12:25 PM - mame (Yusuke Endoh)

- Status changed from Feedback to Closed

I think so. We can keep it unless any serious problem is reported after preview2. Thanks for your quick action!

I'm slightly worried about its very inefficient implementation, but I don't know whether it matters because I understand the use case. Anyway, we can refine it after 2.0.0.

--

Yusuke Endoh mame@tsg.ne.jp