

Ruby - Bug #10239

Regexp.quote() and default encoding

09/14/2014 09:54 AM - shevegen (Robert A. Heiler)

Status:	Closed	Backport: 2.0.0: UNKNOWN, 2.1: UNKNOWN
Priority:	Normal	
Assignee:	zzak (zzak _)	
Target version:		
ruby -v:	ruby 2.1.2p95 (2014-05-08 revision 45877) [x86_64-linux]	
Description		
Hello, I am not sure if this is a bug, or unexpected behaviour (for me). I will simply report it, I am sure you guys know how and if to handle this anyway. I believe it should be documented at least in the official documentation if it is not a bug. The situation is that I have several strings with mixed encodings. Some will have automatically UTF8, some US-ASCII, and yet some others will have ASCII-8BIT. I noticed that Regexp.quote() change the encoding of the string in question in the same project unfortunately, and no way to change that (as some of that gets set from the outside world to me). Here is proof for Regexp.quote() changing the encoding, where x is my test variable - a string: <pre>x = "abc"; x.encoding # => #Encoding:US-ASCII</pre> <pre>x.encode!('ASCII-8BIT'); x.encoding # => #Encoding:ASCII-8BIT</pre> Ok, all works fine, it defaulted to US-ASCII but is not ASCII-8BIT. Next: <pre>test = Regexp.quote(x); test.encoding # => #Encoding:US-ASCII</pre> Suddenly the new string that is returned has another encoding. I looked at the documentation: http://www.ruby-doc.org/core-2.1.2/Regexp.html#method-c-quote But there is no mention that this method would return a new String object with a different encoding. I would have expected it to not change the encoding of the argument-string object there. Perhaps the documentation could mention that it will ignore the original encoding of the string given?		

Associated revisions

Revision 32ec6dd5c7cb89979d48100acf8971ac09e0d02e - 07/22/2019 09:43 PM - jeremyevans (Jeremy Evans)

Document encoding of string returned by Regexp.quote [ci skip]

Also, remove documentation about returning self, which makes no sense as self would be the Regexp class. It could be interpreted as return the argument if no changes were made, but that hasn't been the behavior at least since 1.8.7 (and probably before).

Fixes [Bug #10239]

Revision 32ec6dd5c7cb89979d48100acf8971ac09e0d02e - 07/22/2019 09:43 PM - jeremyevans (Jeremy Evans)

Document encoding of string returned by Regexp.quote [ci skip]

Also, remove documentation about returning self, which makes no sense as self would be the Regexp class. It could be interpreted as return the argument if no changes were made, but that hasn't been the behavior at least since 1.8.7 (and probably before).

Fixes [Bug #10239]

Revision 32ec6dd5 - 07/22/2019 09:43 PM - jeremyevans (Jeremy Evans)

Document encoding of string returned by Regexp.quote [ci skip]

Also, remove documentation about returning self, which makes no sense as self would be the Regexp class. It could be interpreted as return the argument if no changes were made, but that hasn't been the behavior at least since 1.8.7 (and probably before).

Fixes [Bug #10239]

History

#1 - 10/29/2014 08:39 AM - nagachika (Tomoyuki Chikanaga)

- *Category set to doc*
- *Status changed from Open to Assigned*
- *Assignee set to zzak (zzak _)*
- *Target version set to 2.2.0*

I think this is intended behavior.

#2 - 01/05/2018 09:01 PM - naruse (Yui NARUSE)

- *Target version deleted (2.2.0)*

#3 - 07/22/2019 09:49 PM - jeremyevans (Jeremy Evans)

- *Status changed from Assigned to Closed*

Applied in changeset [git|32ec6dd5c7cb89979d48100acf8971ac09e0d02e](https://github.com/ruby/ruby/commit/32ec6dd5c7cb89979d48100acf8971ac09e0d02e).

Document encoding of string returned by Regexp.quote [ci skip]

Also, remove documentation about returning self, which makes no sense as self would be the Regexp class. It could be interpreted as return the argument if no changes were made, but that hasn't been the behavior at least since 1.8.7 (and probably before).

Fixes [Bug [#10239](#)]