

Ruby master - Bug #11495

[Documentation] Please improve documentation for Regexp.new() and clarify the 3 argument call

08/29/2015 11:04 PM - shevegen (Robert A. Heiler)

Status:	Closed	
Priority:	Normal	
Assignee:	zzak (Zachary Scott)	
Target version:		
ruby -v:		Backport: 2.1: DONE, 2.2: DONE, 2.3: DONE
Description		
Hello,		
The documentation at Regepx.new here:		
http://ruby-doc.org/core-2.2.3/Regexp.html#method-c-new		
It shows a lot of information, in particular four ways to call it:		
<pre>new(string, [options [, kcode]]) -> regexp new(regexp) -> regexp compile(string, [options [, kcode]]) -> regexp compile(regexp) -> regexp</pre>		
However, the examples given do not show an example of where 3 arguments are passed.		
Today this came up on IRC where someone asked how to use respectively what the meaning of kcode is, via example.		
Can someone add an example of where/when to use 3 arguments to this method please? Thank you.		

Associated revisions

Revision d9010589 - 02/09/2016 04:54 AM - zzak (Zachary Scott)

- re.c: Remove deprecated kcode argument from Regexp.new and compile patch provided by Dylan Pulliam [Bug #11495]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@53784 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 53784 - 02/09/2016 04:54 AM - zzak (Zachary Scott)

- re.c: Remove deprecated kcode argument from Regexp.new and compile patch provided by Dylan Pulliam [Bug #11495]

Revision 53784 - 02/09/2016 04:54 AM - zzak (Zachary Scott)

- re.c: Remove deprecated kcode argument from Regexp.new and compile patch provided by Dylan Pulliam [Bug #11495]

Revision 53784 - 02/09/2016 04:54 AM - zzak (Zachary Scott)

- re.c: Remove deprecated kcode argument from Regexp.new and compile patch provided by Dylan Pulliam [Bug #11495]

Revision 53784 - 02/09/2016 04:54 AM - zzak (Zachary Scott)

- re.c: Remove deprecated kcode argument from Regexp.new and compile patch provided by Dylan Pulliam [Bug #11495]

Revision b58ee667 - 02/25/2016 11:00 AM - usa (Usaku NAKAMURA)

merge revision(s) 53784: [Backport #11495]

```
* re.c: Remove deprecated kcode argument from Regexp.new and compile
```

patch provided by Dylan Pulliam [Bug #11495]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_2_1@53937 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 53937 - 02/25/2016 11:00 AM - usa (Usaku NAKAMURA)

merge revision(s) 53784: [Backport #11495]

* re.c: Remove deprecated kcode argument from Regexp.new and compile
patch provided by Dylan Pulliam [Bug #11495]

Revision 3e7a554e - 03/09/2016 03:20 PM - nagachika (Tomoyuki Chikanaga)

merge revision(s) 51006,53784: [Backport #11495]

* re.c: Update documentation for Regexp class.
[fix GH-937][ci skip] Patch by @davydovanton

* re.c: Remove deprecated kcode argument from Regexp.new and compile
patch provided by Dylan Pulliam [Bug #11495]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_2_2@54071 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 54071 - 03/09/2016 03:20 PM - nagachika (Tomoyuki Chikanaga)

merge revision(s) 51006,53784: [Backport #11495]

* re.c: Update documentation for Regexp class.
[fix GH-937][ci skip] Patch by @davydovanton

* re.c: Remove deprecated kcode argument from Regexp.new and compile
patch provided by Dylan Pulliam [Bug #11495]

Revision 14e5aab8 - 03/29/2016 05:52 AM - naruse (Yui NARUSE)

merge revision(s) 53784: [Backport #11495]

* re.c: Remove deprecated kcode argument from Regexp.new and compile
patch provided by Dylan Pulliam [Bug #11495]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_2_3@54361 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 54361 - 03/29/2016 05:52 AM - naruse (Yui NARUSE)

merge revision(s) 53784: [Backport #11495]

* re.c: Remove deprecated kcode argument from Regexp.new and compile
patch provided by Dylan Pulliam [Bug #11495]

History

#1 - 08/29/2015 11:05 PM - shevegen (Robert A. Heiler)

The current examples are:

```
r1 = Regexp.new('^a-z+:\s+\w+') #=> /^a-z+:\s+\w+/
r2 = Regexp.new('cat', true) #=> /cat/i
r3 = Regexp.new(r2) #=> /cat/i
r4 = Regexp.new('dog', Regexp::EXTENDED | Regexp::IGNORECASE) #=> /dog/ix
```

As you can see, we have examples for one argument, and two arguments given but not for three arguments.

Thanks!

#2 - 10/12/2015 12:10 AM - dpulliam (Dylan Pulliam)

- File 0001-Improve-documentation-for-Regexp.new.patch added

Robert A. Heiler wrote:

The current examples are:

```
r1 = Regexp.new('^a-z+:\s+\w+') #=> /^a-z+:\s+\w+/
r2 = Regexp.new('cat', true) #=> /cat/i
```

```
r3 = Regexp.new(r2)           #=> /cat/i
r4 = Regexp.new('dog', Regexp::EXTENDED | Regexp::IGNORECASE) #=> /dog/ix
```

As you can see, we have examples for one argument, and two arguments given but not for three arguments.

Thanks!

Hey Robert,

TL;DR

\$KCODE is deprecated and does not affect the use of Regexp at this point. I have a patch ready to go that removes [, kcode] from the doc. That being said I would be happy to modify the patch and just submit an example of using Regexp with 3 parameters.

```
r5 = Regexp.new('fish', Regexp::IGNORECASE, 'utf8')
```

The Good Stuff

The kcode parameter in new(string, [options [, kcode]]) or compile(string, [options [, kcode]]) is used to set the global \$KCODE variable which basically sets the encoding to be used. This could be UTF-8, Unicode, or any other valid encoding etc.

I believe the example you are looking for would be the following:

```
r5 = Regexp.new('fish', Regexp::IGNORECASE, 'utf8')
```

This example is not dependent on the [option] Regexp::IGNORECASE. Any other accepted option should still work with the above example.

However, \$KCODE was deprecated in Ruby 1.9 so you will find that setting the encoding does not do anything and you will receive a warning in most cases. You can see this here:

```
ruby regexp_example.rb
regexp_example.rb:3: warning: encoding option is ignored - utf8
```

or

```
irb(main):001:0> Regexp.new('fish', Regexp::IGNORECASE, 'utf8')
(irb):1: warning: encoding option is ignored - utf8
=> /fish/i
```

However setting the encoding at the console does not provide a warning:

```
ruby -KU #=> Runs Ruby with UTF-8 encoding
```

I went ahead and prepared a patch for the documentation. You should be able to find it attached to this comment. I think the best course of action would be to remove [, kcode] since you cannot pass any option that will actually change the outcome. The internal C code will default to 'n' or 'N' which you can pass in without warning but basically just says "don't specify encoding just treat as binary string". That being said since it was deprecated I would be happy to modify the patch and just add the example for now.

If you would like more information on regular expressions and kcodes in ruby go ahead and checkout the following links. They helped a lot when I was researching this topic.

<http://nuclearsquid.com/writings/ruby-1-9-what-s-new-what-s-changed/>
<http://graysoftinc.com/character-encodings/the-kcode-variable-and-jcode-library>
<https://www.ruby-forum.com/topic/178589>
<http://yehudakatz.com/2010/05/05/ruby-1-9-encodings-a-primer-and-the-solution-for-rails/>

#3 - 01/14/2016 05:44 PM - dpulliam (Dylan Pulliam)

- Tracker changed from Misc to Bug
- Assignee set to zzak (Zachary Scott)
- Backport set to 2.0.0: UNKNOWN, 2.1: UNKNOWN, 2.2: UNKNOWN, 2.3: UNKNOWN

#4 - 02/09/2016 04:54 AM - zzak (Zachary Scott)

- Status changed from Open to Closed

Applied in changeset r53784.

-
- re.c: Remove deprecated kcode argument from Regexp.new and compile patch provided by Dylan Pulliam [Bug #11495]

#5 - 02/23/2016 08:53 AM - usa (Usaku NAKAMURA)

- Backport changed from 2.0.0: UNKNOWN, 2.1: UNKNOWN, 2.2: UNKNOWN, 2.3: UNKNOWN to 2.1: REQUIRED, 2.2: REQUIRED, 2.3: REQUIRED

#6 - 02/25/2016 11:00 AM - usa (Usaku NAKAMURA)

- Backport changed from 2.1: REQUIRED, 2.2: REQUIRED, 2.3: REQUIRED to 2.1: DONE, 2.2: REQUIRED, 2.3: REQUIRED

ruby_2_1 r53937 merged revision(s) 53784.

#7 - 03/09/2016 03:21 PM - nagachika (Tomoyuki Chikanaga)

- Backport changed from 2.1: DONE, 2.2: REQUIRED, 2.3: REQUIRED to 2.1: DONE, 2.2: DONE, 2.3: REQUIRED

r51006 and r53784 were backported into ruby_2_2 branch at r54071.

#8 - 03/29/2016 05:57 AM - naruse (Yui NARUSE)

- Backport changed from 2.1: DONE, 2.2: DONE, 2.3: REQUIRED to 2.1: DONE, 2.2: DONE, 2.3: DONE

ruby_2_3 r54361 merged revision(s) 53784.

Files

0001-Improve-documentation-for-Regexp.new.patch	1.55 KB	10/12/2015	dpulliam (Dylan Pulliam)
---	---------	------------	--------------------------