

## Backport193 - Backport #8210

### Multibyte character interfering with end-line character within a regex

04/03/2013 03:22 PM - sawa (Tsuyoshi Sawada)

<b>Status:</b>	Closed
<b>Priority:</b>	Normal
<b>Assignee:</b>	usa (Usaku NAKAMURA)
<b>Description</b>	
=begin With this regex:  regex1 = /\z/  the following strings match as expected:  "hello" =~ regex1 # => 5 "##### " =~ regex1 # => 5  but with these regexes:  regex2 = /#\$/?\z/ regex3 = /\n?\z/  they show difference:  "hello" =~ regex2 # => 5 "hello" =~ regex3 # => 5 "##### " =~ regex2 # => nil "##### " =~ regex3 # => nil  The string encoding is UTF-8, and the OS is Linux (i.e., \$/ is "\n"). I expect them to behave the same, and believe this is a bug. =end	

#### Associated revisions

##### Revision 536a3274 - 04/13/2013 11:30 AM - naruse (Yui NARUSE)

- Merge Onigmo 5.13.4 f22cf2e566712cace60d17f84d63119d7c5764ee. [bug] fix problem with optimization of \z (Issue #16) [Bug #8210]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@40276 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

##### Revision 40276 - 04/13/2013 11:30 AM - naruse (Yui NARUSE)

- Merge Onigmo 5.13.4 f22cf2e566712cace60d17f84d63119d7c5764ee. [bug] fix problem with optimization of \z (Issue #16) [Bug #8210]

##### Revision 40276 - 04/13/2013 11:30 AM - naruse (Yui NARUSE)

- Merge Onigmo 5.13.4 f22cf2e566712cace60d17f84d63119d7c5764ee. [bug] fix problem with optimization of \z (Issue #16) [Bug #8210]

##### Revision 40276 - 04/13/2013 11:30 AM - naruse (Yui NARUSE)

- Merge Onigmo 5.13.4 f22cf2e566712cace60d17f84d63119d7c5764ee. [bug] fix problem with optimization of \z (Issue #16) [Bug #8210]

##### Revision 40276 - 04/13/2013 11:30 AM - naruse (Yui NARUSE)

- Merge Onigmo 5.13.4 f22cf2e566712cace60d17f84d63119d7c5764ee. [bug] fix problem with optimization of \z (Issue #16) [Bug #8210]

##### Revision 40276 - 04/13/2013 11:30 AM - naruse (Yui NARUSE)

- Merge Onigmo 5.13.4 f22cf2e566712cace60d17f84d63119d7c5764ee. [bug] fix problem with optimization of \z (Issue #16) [Bug #8210]

## Revision 40276 - 04/13/2013 11:30 AM - naruse (Yui NARUSE)

- Merge Onigmo 5.13.4 f22cf2e566712cace60d17f84d63119d7c5764ee. [bug] fix problem with optimization of \z (Issue #16) [Bug #8210]

## Revision 8c43fc02 - 04/19/2013 04:40 PM - nagachika (Tomoyuki Chikanaga)

merge revision(s) 40276: [Backport #8210]

```
* Merge Onigmo 5.13.4 f22cf2e566712cace60d17f84d63119d7c5764ee.  
[bug] fix problem with optimization of \z (Issue #16) [Bug #8210]
```

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby\_2\_0\_0@40384 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

## Revision e8905e10 - 05/14/2013 01:41 AM - usa (Usaku NAKAMURA)

- regex.c (onig\_search): fix problem with optimization of \z. [Backport #8210] patched by k\_tanaka at [ruby-core:54251].

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby\_1\_9\_3@40713 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

## Revision 40713 - 05/14/2013 01:41 AM - usa (Usaku NAKAMURA)

- regex.c (onig\_search): fix problem with optimization of \z. [Backport #8210] patched by k\_tanaka at [ruby-core:54251].

## History

---

### #1 - 04/03/2013 03:27 PM - sawa (Tsuyoshi Sawada)

```
=begin  
A different regex:  
  
regex4 = /[[:space:]]?\z/  
  
seems to work as expected:  
  
"hello" =~ regex4 # => 5  
"    " =~ regex4 # => 5  
  
=end
```

### #2 - 04/03/2013 03:37 PM - sawa (Tsuyoshi Sawada)

```
=begin  
Still a different regex:  
  
regex5 = /\n?$/  
  
seems to work as expected:  
  
"hello" =~ regex5 # => 5  
"    " =~ regex5 # => 5  
  
=end
```

### #3 - 04/06/2013 09:58 AM - sawa (Tsuyoshi Sawada)

```
=begin  
The problem seems to happen with combination of a certain token, ?, and \z.  
  
"    " =~ /a?\z/ # => nil  
"    " =~ / ?\z/ # => nil  
"    " =~ /\t?\z/ # => nil  
"    " =~ /\n?\z/ # => nil  
"    " =~ /\s?\z/ # => nil  
"    " =~ /.?\z/ # => 4  
"    " =~ /\S?\z/ # => 4  
"    " =~ /\W?\z/ # => 5  
"    " =~ /\_?\z/ # => 5  
"    " =~ /\w?\z/ # => 5  
  
=end
```

### #4 - 04/06/2013 07:34 PM - sawa (Tsuyoshi Sawada)

Is this bug report wrong? If so, please note so.

**#5 - 04/06/2013 08:15 PM - naruse (Yui NARUSE)**

- *Category set to M17N*
- *Status changed from Open to Assigned*
- *Assignee set to naruse (Yui NARUSE)*
- *Target version set to 2.1.0*

sawa (Tsuyoshi Sawada) wrote:

Is this bug report wrong? If so, please note so.

This looks really bug of oniguruma/onigmo.

**#6 - 04/09/2013 04:05 AM - acheong87 (Andrew Cheong)**

Contributing notes regarding this bug can be found here: <http://stackoverflow.com/a/15885857/925913>.

**#7 - 04/09/2013 02:42 PM - rondinif (Franco Rondini)**

Just edited the [answer](#) and [test code available](#)

**#8 - 04/10/2013 12:41 AM - k\_takata (Ken Takata)**

- *File fix-8210-1.diff added*
- *File fix-8210-2.diff added*

This problem was caused by optimization of \z.  
I wrote two patches to fix this problem.

Maybe fix-8210-1.diff is more efficient than fix-8210-2.diff,  
but the former one tries to do backward search when 'start==range'  
after 'start' is adjusted. This behavior is a little bit confusing.

**#9 - 04/11/2013 12:46 PM - sawa (Tsuyoshi Sawada)**

Is either of k\_takata's bug fix going to be incorporated?

**#10 - 04/11/2013 11:31 PM - naruse (Yui NARUSE)**

k\_takata (Ken Takata) wrote:

This problem was caused by optimization of \z.  
I wrote two patches to fix this problem.

Maybe fix-8210-1.diff is more efficient than fix-8210-2.diff,  
but the former one tries to do backward search when 'start==range'  
after 'start' is adjusted. This behavior is a little bit confusing.

k\_takata (Ken Takata) wrote:

This problem was caused by optimization of \z.  
I wrote two patches to fix this problem.

Maybe fix-8210-1.diff is more efficient than fix-8210-2.diff,  
but the former one tries to do backward search when 'start==range'  
after 'start' is adjusted. This behavior is a little bit confusing.

I think -1 is suitable because it looks to keep original intention more than -2.

**#11 - 04/13/2013 07:31 PM - k\_takata (Ken Takata)**

- *File fix-8210-1-update.diff added*

I think -1 is suitable because it looks to keep original intention more than -2.

Thanks for your comment.

I have updated onigmo's tmp/ruby-2.0.x branch.

<https://github.com/k-takata/Onigmo/tree/f22cf2e566712cace60d17f84d63119d7c5764ee>

I also attach an updated patch so that can be applied to Ruby 1.9.3.

**#12 - 04/13/2013 08:30 PM - naruse (Yui NARUSE)**

- Status changed from Assigned to Closed

- % Done changed from 0 to 100

This issue was solved with changeset r40276.

Tsuyoshi, thank you for reporting this issue.

Your contribution to Ruby is greatly appreciated.

May Ruby be with you.

- 
- Merge Onigmo 5.13.4 f22cf2e566712cace60d17f84d63119d7c5764ee. [bug] fix problem with optimization of \z (Issue [#16](#)) [Bug [#8210](#)]

**#13 - 04/13/2013 08:32 PM - naruse (Yui NARUSE)**

- Tracker changed from Bug to Backport

- Project changed from Ruby trunk to Backport200

- Category deleted (M17N)

- Status changed from Closed to Assigned

- Assignee changed from naruse (Yui NARUSE) to nagachika (Tomoyuki Chikanaga)

- Target version deleted (2.1.0)

**#14 - 04/13/2013 10:17 PM - k\_takata (Ken Takata)**

I think it's better to backport this patch to Ruby 1.9.3 too.

**#15 - 04/20/2013 01:40 AM - nagachika (Tomoyuki Chikanaga)**

- Status changed from Assigned to Closed

This issue was solved with changeset r40384.

Tsuyoshi, thank you for reporting this issue.

Your contribution to Ruby is greatly appreciated.

May Ruby be with you.

---

merge revision(s) 40276: [Backport [#8210](#)]

```
* Merge Onigmo 5.13.4 f22cf2e566712cace60d17f84d63119d7c5764ee.  
[bug] fix problem with optimization of \z (Issue #16) [Bug #8210]
```

**#16 - 04/20/2013 01:45 AM - nagachika (Tomoyuki Chikanaga)**

- Project changed from Backport200 to Backport193

- Status changed from Closed to Assigned

- Assignee changed from nagachika (Tomoyuki Chikanaga) to usa (Usaku NAKAMURA)

Move to Backport93.

But Onigmo is merged after 2.0. I didn't confirm this patch can merge to ruby\_1\_9\_3...

**#17 - 05/14/2013 10:41 AM - usa (Usaku NAKAMURA)**

- Status changed from Assigned to Closed

This issue was solved with changeset [r40713](#).

Tsuyoshi, thank you for reporting this issue.

Your contribution to Ruby is greatly appreciated.

May Ruby be with you.

- 
- regex.c (onig\_search): fix problem with optimization of \z. [Backport [#8210](#)] patched by k\_tanaka at [ruby-core:54251].

## #18 - 05/14/2013 06:23 PM - k\_takata (Ken Takata)

Hi usa,

- regexec.c (onig\_search): fix problem with optimization of \z. [Backport [#8210](#)] patched by k\_tanaka at [ruby-core:54251].

Thank you for merging my patch.

BTW, my name is not tanaka...

### Files

---

fix-8210-1.diff	742 Bytes	04/10/2013	k_takata (Ken Takata)
fix-8210-2.diff	491 Bytes	04/10/2013	k_takata (Ken Takata)
fix-8210-1-update.diff	834 Bytes	04/13/2013	k_takata (Ken Takata)