

Ruby trunk - Bug #14363

each_grapheme_cluster.size returns the wrong size

01/16/2018 07:52 AM - sos4nt (Stefan Schüßler)

Status: Closed	
Priority: Normal	
Assignee:	
Target version:	
ruby -v: ruby 2.5.0p0 (2017-12-25 revision 61468) [x86_64-darwin15]	Backport: 2.3: DONTNEED, 2.4: DONTNEED, 2.5: DONE
Description	
Ruby 2.5 adds String#each_grapheme_cluster to enumerate the string's grapheme clusters:	
<pre>str = "a\u0300i\u0301" #=> "àï" str.each_grapheme_cluster.to_a #=> ["à", "ï"]</pre>	
Unfortunately, the enumerator's size doesn't work as expected:	
<pre>str.each_grapheme_cluster.size #=> 4</pre>	
The source code reveals that it invokes rb_str_each_char_size, so it is equivalent to each_char.size:	
<pre>static VALUE rb_str_each_grapheme_cluster(VALUE str) { RETURN_SIZED_ENUMERATOR(str, 0, 0, rb_str_each_char_size); return rb_str_enumerate_grapheme_clusters(str, 0); }</pre>	
If the grapheme enumerator's size cannot be calculated lazily, each_grapheme_cluster.size should return nil to indicate that.	

Associated revisions

Revision 613decd0 - 03/22/2018 06:59 AM - naruse (Yui NARUSE)

each_grapheme_cluster shouldn't return size [Bug #14363]

From: Stefan Schüßler mail@stefanschuessler.de

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@62888 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 62888 - 03/22/2018 06:59 AM - naruse (Yui NARUSE)

each_grapheme_cluster shouldn't return size [Bug #14363]

From: Stefan Schüßler mail@stefanschuessler.de

Revision 62888 - 03/22/2018 06:59 AM - naruse (Yui NARUSE)

each_grapheme_cluster shouldn't return size [Bug #14363]

From: Stefan Schüßler mail@stefanschuessler.de

Revision 6e0f5b84 - 03/22/2018 07:58 AM - naruse (Yui NARUSE)

Revert "each_grapheme_cluster shouldn't return size [Bug #14363]"

This reverts commit r62887.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@62891 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 62891 - 03/22/2018 07:58 AM - naruse (Yui NARUSE)

Revert "each_grapheme_cluster shouldn't return size [Bug #14363]"

This reverts commit r62887.

Revision 62891 - 03/22/2018 07:58 AM - naruse (Yui NARUSE)

Revert "each_grapheme_cluster shouldn't return size [Bug #14363]"

This reverts commit r62887.

Revision 41b2ef46 - 03/22/2018 07:58 AM - naruse (Yui NARUSE)

fix each_grapheme_cluster's size [Bug #14363]

From: Hugo Peixoto hugo.peixoto@gmail.com

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@62892 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 62892 - 03/22/2018 07:58 AM - naruse (Yui NARUSE)

fix each_grapheme_cluster's size [Bug #14363]

From: Hugo Peixoto hugo.peixoto@gmail.com

Revision 62892 - 03/22/2018 07:58 AM - naruse (Yui NARUSE)

fix each_grapheme_cluster's size [Bug #14363]

From: Hugo Peixoto hugo.peixoto@gmail.com

Revision c40df5a7 - 03/22/2018 11:18 AM - naruse (Yui NARUSE)

merge revision(s) 62892,62893: [Backport #14363]

```
fix each_grapheme_cluster's size [Bug #14363]
```

```
From: Hugo Peixoto <hugo.peixoto@gmail.com>
```

```
Factor out get_reg_grapheme_cluster
```

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_2_5@62896 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 62896 - 03/22/2018 11:18 AM - naruse (Yui NARUSE)

merge revision(s) 62892,62893: [Backport #14363]

```
fix each_grapheme_cluster's size [Bug #14363]
```

```
From: Hugo Peixoto <hugo.peixoto@gmail.com>
```

```
Factor out get_reg_grapheme_cluster
```

History

#1 - 03/21/2018 04:17 PM - hugopeixoto (Hugo Peixoto)

- File *each_grapheme_cluster_size_real.patch* added

- File *each_grapheme_cluster_size_nil.patch* added

Calculating the enumerator size here requires iterating through the whole text and do grapheme detection on all bytes, so I'm not sure what's the right approach.

I'm attaching two patches, one that makes it return nil and one that does the actual count. Both patches have tests attached.

#2 - 03/22/2018 06:59 AM - naruse (Yui NARUSE)

- Status changed from *Open* to *Closed*

Applied in changeset [trunk|r62892](#).

```
fix each_grapheme_cluster's size [Bug #14363]
```

From: Hugo Peixoto hugo.peixoto@gmail.com

#3 - 03/22/2018 07:59 AM - naruse (Yui NARUSE)

- Backport changed from 2.3: UNKNOWN, 2.4: UNKNOWN, 2.5: UNKNOWN to 2.3: DONTNEED, 2.4: DONTNEED, 2.5: REQUIRED

#4 - 03/22/2018 11:18 AM - naruse (Yui NARUSE)

- Backport changed from 2.3: *DONTNEED*, 2.4: *DONTNEED*, 2.5: *REQUIRED* to 2.3: *DONTNEED*, 2.4: *DONTNEED*, 2.5: *DONE*

ruby_2_5 r62896 merged revision(s) 62892,62893.

Files

each_grapheme_cluster_size_nil.patch	921 Bytes	03/21/2018	hugopeixoto (Hugo Peixoto)
each_grapheme_cluster_size_real.patch	3.03 KB	03/21/2018	hugopeixoto (Hugo Peixoto)