

Backport200 - Backport #8165

Problems with require

03/26/2013 09:34 PM - Krugloff (Alexandr Kruglov)

Status:	Closed
Priority:	Normal
Assignee:	nagachika (Tomoyuki Chikanaga)
Description	
Require doesn't work correctly if path contains cyrillic.	
require '/home/mak/test.rb' #-> true	
require '/home/mak/test.rb' #-> false	
require '/home/mak/Проекты/test.rb' #-> true	
require '/home/mak/Проекты/test.rb' #-> true	

Associated revisions

Revision 07636ed4 - 04/05/2013 03:30 PM - shirosaki

- load.c (features_index_add): use rb_str_subseq() to specify C string position properly to fix require non ascii path.
[ruby-core:53733] [Bug #8165]
- test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
a test for the above.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@40135 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 40135 - 04/05/2013 03:30 PM - shirosaki

- load.c (features_index_add): use rb_str_subseq() to specify C string position properly to fix require non ascii path.
[ruby-core:53733] [Bug #8165]
- test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
a test for the above.

Revision 40135 - 04/05/2013 03:30 PM - shirosaki

- load.c (features_index_add): use rb_str_subseq() to specify C string position properly to fix require non ascii path.
[ruby-core:53733] [Bug #8165]
- test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
a test for the above.

Revision 40135 - 04/05/2013 03:30 PM - shirosaki

- load.c (features_index_add): use rb_str_subseq() to specify C string position properly to fix require non ascii path.
[ruby-core:53733] [Bug #8165]

- test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
a test for the above.

Revision 40135 - 04/05/2013 03:30 PM - shirosaki

- load.c (features_index_add): use rb_str_subseq() to specify C string position properly to fix require non ascii path.
[ruby-core:53733] [Bug #8165]
- test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
a test for the above.

Revision 40135 - 04/05/2013 03:30 PM - shirosaki

- load.c (features_index_add): use rb_str_subseq() to specify C string position properly to fix require non ascii path.
[ruby-core:53733] [Bug #8165]
- test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
a test for the above.

Revision 40135 - 04/05/2013 03:30 PM - shirosaki

- load.c (features_index_add): use rb_str_subseq() to specify C string position properly to fix require non ascii path.
[ruby-core:53733] [Bug #8165]
- test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
a test for the above.

Revision 88d0a18c - 04/20/2013 03:14 PM - nagachika (Tomoyuki Chikanaga)

merge revision(s) 40135,40148,40173: [Backport #8165]

- * load.c (features_index_add): use rb_str_subseq() to specify C string position properly to fix require non ascii path.
[ruby-core:53733] [Bug #8165]
- * test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
a test for the above.
- * test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
fix load path for encoding to run the test as stand-alone.
- * test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
RUBY_PLATFORM should escape as Regexp,
because RUBY_PLATFORM may contain '.'.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_2_0_0@40398 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 40398 - 04/20/2013 03:14 PM - nagachika (Tomoyuki Chikanaga)

merge revision(s) 40135,40148,40173: [Backport #8165]

- * load.c (features_index_add): use rb_str_subseq() to specify C string position properly to fix require non ascii path.
[ruby-core:53733] [Bug #8165]

```
* test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
  a test for the above.

* test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
  fix load path for encoding to run the test as stand-alone.

* test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):
  RUBY_PLATFORM should escape as Regexp,
  because RUBY_PLATFORM may contain '.'.
```

History

#1 - 03/26/2013 09:56 PM - luislavena (Luis Lavena)

- Status changed from Open to Feedback

Hello,

Please provide more details of the platform/OS, version and filesystem used.

Thank you.

#2 - 03/26/2013 10:17 PM - mame (Yusuke Endoh)

- Status changed from Feedback to Assigned

- Assignee set to tarui (Masaya Tarui)

- Target version set to 2.1.0

It reproduces on my Ubuntu 12.10 with ext4 filesystem.

```
$ cat t.rb
```

coding: UTF-8

```
p require("/home/mame/????/test.rb")
p require("/home/mame/????/test.rb")
p require("/home/mame/????/test.rb")
p(*$)
```

```
$ ruby -v t.rb
ruby 2.0.0p0 (2013-02-24 revision 39474) [x86_64-linux]
true
true
true
"enumerator.so"
snip
"/home/mame/????/test.rb"
"/home/mame/????/test.rb"
"/home/mame/????/test.rb"
```

```
$ ruby-1.9.3-p392 -v t.rb
ruby 1.9.3p392 (2013-02-22 revision 39386) [x86_64-linux]
true
false
false
"enumerator.so"
snip
"/home/mame/????/test.rb"
```

I guess this is related to the performance improvement of require in 2.0.0.
Tarui-san or Shirotsuki-san, do you have any idea?

--

Yusuke Endoh mame@tsg.ne.jp

#3 - 03/27/2013 02:20 AM - funny_falcon (Yura Sokolov)

This is occasionally fixed in trunk by <https://bugs.ruby-lang.org/issues/8048>,
by de-abusing Ruby Strings and using plain C strings.

Also backported version of <https://bugs.ruby-lang.org/issues/8158> fixes this too,
cause it removes string comparison at all.

#4 - 03/28/2013 12:49 AM - h.shirosaki (Hiroshi Shirosaki)

It seems `short_feature` is not proper if the path contains non-ascii characters.
`rb_str_substr()` would look up string position per character, not per byte. `rb_str_subseq()` seems to work for non-ascii.

```
diff --git a/load.c b/load.c
index 7c69461..dab493d 100644
--- a/load.c
+++ b/load.c
@@ -247,16 +247,16 @@ features_index_add(VALUE feature, VALUE offset)
if (p < feature_str)
break;
/* Now *p == '/'. We reach this point for every '/' in feature. */

• short_feature = rb_str_substr(feature, p + 1 - feature_str, feature_end - p - 1);
• short_feature = rb_str_subseq(feature, p + 1 - feature_str, feature_end - p - 1); features_index_add_single(short_feature, offset); if (ext) {
• short_feature = rb_str_substr(feature, p + 1 - feature_str, ext - p - 1);
• short_feature = rb_str_subseq(feature, p + 1 - feature_str, ext - p - 1); features_index_add_single(short_feature, offset); } }
features_index_add_single(feature, offset); if (ext) {
• short_feature = rb_str_substr(feature, 0, ext - feature_str);
• short_feature = rb_str_subseq(feature, 0, ext - feature_str); features_index_add_single(short_feature, offset); } }
```

#5 - 03/29/2013 07:26 PM - h.shirosaki (Hiroshi Shirosaki)

- File 0001-load.c-fix-require-with-non-ascii-path.patch added

One more issue seems string encoding.
When looking up a feature name in `rb_feature_p()`, encoding information lacks.
So `short_feature` should not have encoding.
If encoding of the two non-ascii strings is different, hash value would be different.

```
load.c: rb_feature_p()

feature_val = rb_str_new(feature, len); // not have encoding

this_feature_index = rb_hash_lookup(features_index, feature_val);
```

I've attached a patch. Tested on both trunk and 2.0.0 branch.

ruby 2.1.0dev (2013-03-29 trunk 39991) [x86_64-linux]
ruby 2.0.0p100 (2013-03-27 revision 39954) [x86_64-linux]

#6 - 03/29/2013 08:53 PM - funny_falcon (Yura Sokolov)

May be it will be simpler to accept <https://bugs.ruby-lang.org/issues/8158> ?
It has less memory requirement, avoids string comparisons, etc

(yes, i'm annoying a bit, sorry)

2013/3/29 h.shirosaki (Hiroshi Shirosaki) h.shirosaki@gmail.com

Issue [#8165](#) has been updated by h.shirosaki (Hiroshi Shirosaki).

File 0001-load.c-fix-require-with-non-ascii-path.patch added

One more issue seems string encoding.
When looking up a feature name in `rb_feature_p()`, encoding information lacks.
So `short_feature` should not have encoding.
If encoding of the two non-ascii strings is different, hash value would be different.

```
load.c: rb_feature_p()

feature_val = rb_str_new(feature, len); // not have encoding

this_feature_index = rb_hash_lookup(features_index, feature_val);
```

I've attached a patch. Tested on both trunk and 2.0.0 branch.

ruby 2.1.0dev (2013-03-29 trunk 39991) [x86_64-linux]

ruby 2.0.0p100 (2013-03-27 revision 39954) [x86_64-linux]

Bug [#8165](https://bugs.ruby-lang.org/issues/8165#change-38028): Problems with require
<https://bugs.ruby-lang.org/issues/8165#change-38028>

Author: Krugloff (Alexandr Kruglov)
Status: Assigned
Priority: Normal
Assignee: tarui (Masaya Tarui)
Category:
Target version: current: 2.1.0
ruby -v: ruby 2.0.0-p0

Require doesn't work correctly if path contains cyrillic.

```
require '/home/mak/test.rb' #-> true  
require '/home/mak/test.rb' #-> false
```

```
require '/home/mak/Проекты/test.rb' #-> true  
require '/home/mak/Проекты/test.rb' #-> true
```

--
<http://bugs.ruby-lang.org/>

#7 - 03/30/2013 01:14 AM - h.shirosaki (Hiroshi Shirosaki)

funny_falcon (Yura Sokolov) wrote:

May be it will be simpler to accept <https://bugs.ruby-lang.org/issues/8158> ?
It has less memory requirement, avoids string comparisons, etc

Indeed your patch would not have this bug since using C string instead of string object.
I think smaller patch or splitting patch might be better to be reviewed.

#8 - 03/30/2013 05:53 AM - funny_falcon (Yura Sokolov)

My patch doesn't store any string value, only hashes of them. It relies on loaded_feature_path for string comparison.

Nor it uses ruby arrays for index storage - it organize them in linked lists which are stored in a single C array.
29.03.2013 20:14 пользователь "h.shirosaki (Hiroshi Shirosaki)" <h.shirosaki@gmail.com> написал:

Issue [#8165](https://bugs.ruby-lang.org/issues/8165) has been updated by h.shirosaki (Hiroshi Shirosaki).

funny_falcon (Yura Sokolov) wrote:

May be it will be simpler to accept
<https://bugs.ruby-lang.org/issues/8158> ?
It has less memory requirement, avoids string comparisons, etc

Indeed your patch would not have this bug since using C string instead of string object.
I think smaller patch or splitting patch might be better to be reviewed.

Bug [#8165](https://bugs.ruby-lang.org/issues/8165#change-38032): Problems with require
<https://bugs.ruby-lang.org/issues/8165#change-38032>

Author: Krugloff (Alexandr Kruglov)
Status: Assigned
Priority: Normal
Assignee: tarui (Masaya Tarui)
Category:
Target version: current: 2.1.0
ruby -v: ruby 2.0.0-p0

Require doesn't work correctly if path contains cyrillic.

```
require '/home/mak/test.rb' #-> true  
require '/home/mak/test.rb' #-> false
```

```
require '/home/mak/Проекты/test.rb' #-> true
require '/home/mak/Проекты/test.rb' #-> true
```

--
<http://bugs.ruby-lang.org/>

#9 - 04/06/2013 12:30 AM - Anonymous

- Status changed from Assigned to Closed
- % Done changed from 0 to 100

This issue was solved with changeset [r40135](#).
Alexandr, thank you for reporting this issue.
Your contribution to Ruby is greatly appreciated.
May Ruby be with you.

-
- `load.c` (`features_index_add`): use `rb_str_subseq()` to specify C string position properly to fix `require non ascii path`.
[[ruby-core:53733](#)] [Bug [#8165](#)]
 - `test/ruby/test_require.rb` (`TestRequire#test_require_nonascii_path`):
a test for the above.

#10 - 04/07/2013 12:10 AM - nagachika (Tomoyuki Chikanaga)

- Tracker changed from Bug to Backport
- Project changed from Ruby master to Backport200
- Status changed from Closed to Assigned
- Assignee changed from *tarui (Masaya Tarui)* to *nagachika (Tomoyuki Chikanaga)*
- Target version deleted (2.1.0)

#11 - 04/16/2013 12:26 AM - nagachika (Tomoyuki Chikanaga)

working memo: backport 40135,40148,40173 but `ruby/test_require.rb TestRequire#test_require_nonascii_path` fails.

#12 - 04/16/2013 06:50 AM - h.shirosaki (Hiroshi Shirosaki)

nagachika (Tomoyuki Chikanaga) wrote:

working memo: backport 40135,40148,40173 but `ruby/test_require.rb TestRequire#test_require_nonascii_path` fails.

Thank you for backport. [#8048](#) would also be needed for this issue.
Hash lookup of `feature_index` fails due to String encoding of the key. [r39874](#) changes hash key from ruby String to C string.

#13 - 04/16/2013 09:46 AM - nagachika (Tomoyuki Chikanaga)

shirosaki san
Thank you for your advice! I'll try to backport with [r39874](#) later.

#14 - 04/21/2013 12:14 AM - nagachika (Tomoyuki Chikanaga)

- Status changed from Assigned to Closed

This issue was solved with changeset [r40398](#).
Alexandr, thank you for reporting this issue.
Your contribution to Ruby is greatly appreciated.
May Ruby be with you.

merge revision(s) 40135,40148,40173: [Backport [#8165](#)]

```
* load.c (features_index_add): use rb_str_subseq() to specify C string
  position properly to fix require non ascii path.
  [ruby-core:53733] [Bug #8165]
```

```
* test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):  
  a test for the above.  
  
* test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):  
  fix load path for encoding to run the test as stand-alone.  
  
* test/ruby/test_require.rb (TestRequire#test_require_nonascii_path):  
  RUBY_PLATFORM should escape as Regexp,  
  because RUBY_PLATFORM may contain '.'.
```

Files

0001-load.c-fix-require-with-non-ascii-path.patch	3.43 KB	03/29/2013	h.shirosaki (Hiroshi Shirosaki)
---------------------------------------------------	---------	------------	---------------------------------