Ruby master - Bug #3376
russian support
06/01/2010 07:58 PM - dcab (Yury Korolev)

<table>
<thead>
<tr>
<th>Status:</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority:</td>
<td>Normal</td>
</tr>
<tr>
<td>Assignee:</td>
<td>naruse (Yui NARUSE)</td>
</tr>
<tr>
<td>Target version:</td>
<td>1.9.2</td>
</tr>
<tr>
<td>ruby -v:</td>
<td>ruby 1.9.2dev (2010-05-31 revision 28117) [x86_64-darwin10.3.0]</td>
</tr>
</tbody>
</table>

Description

=begin

1. String#upcase/String#downcase don't work with russian characters

ruby-1.9.2-preview3 > "привет".upcase
=> "привет"
ruby-1.9.2-preview3 > "ПРИВЕТ".downcase
=> "ПРИВЕТ"

1. Regex doesn't work with POSIX bracket classes

ruby-1.9.2-preview3 > "привет" =~ /[:alpha:]/
=> nil
It's working with unicode groups

ruby-1.9.2-preview3 > "привет" =~ /\p{L}/
=> 0

1. Negated unicode groups don't work

SyntaxError: (irb):13: invalid character property name {L}: /\P{L}/
from /Users/yury/.rvm/rubies/ruby-1.9.2-preview3/bin/irb:17:in `'

=end

Related issues:
Related to CommonRuby - Feature #10085: Add non-ASCII case conversion to Stri... Closed

History

#1 - 06/01/2010 08:33 PM - naruse (Yui NARUSE)
=begin

1. String#upcase/String#downcase don't work with russian characters

Ruby's String#upcase is for ASCII.

1. Regex doesn't work with POSIX bracket classes "привет" =~ /[:alpha:]/

"привет" =~ /alpha/ is correct.

1. Negated unicode groups don't work

It seems a bug. I'll check it, thanks.
=end

#2 - 06/01/2010 08:34 PM - naruse (Yui NARUSE)
- Assignee set to naruse (Yui NARUSE)
Hi,
Thank you for your report.

2010/6/1 Yury Korolev redmine@ruby-lang.org:

1. String#upcase/String#downcase don't work with russian characters

This is the spec. Traditionally, they have handled only ASCII characters.

1. Regex doesn't work with POSIX bracket classes

```
ruby-1.9.2-preview3 > "привет" =~ /[:alpha:]/
nil
```

This is also the spec. This spec was changed after 1.9.1.
See NEWS: http://svn.ruby-lang.org/repos/ruby/tags/v1_9_2_preview3/NEWS

The rationale is that many people expects \d and \s to match ASCII only.

1. Negated unicode groups don't work

```
SyntaxError: (irb):13: invalid character property name \{L\}: /\P{L}/
```

This looks like a bug. We'll invetigave it. Thanks!

--
Yusuke Endoh mame@tsg.ne.jp

This issue was solved with changeset r28120.

Yury, thank you for reporting this issue.
Your contribution to Ruby is greatly appreciated.
May Ruby be with you.

#4 - 06/01/2010 11:27 PM - naruse (Yui NARUSE)
- Status changed from Open to Closed
- % Done changed from 0 to 100

#5 - 07/23/2014 11:06 AM - duerst (Martin Dürst)
- Related to Feature #10085: Add non-ASCII case conversion to String#upcase/downcase/swapcase/capitalize added