Ruby master - Bug #7092
StringScanner start-of-line matches false positives, while lookbehind for a newline gives false negatives

10/01/2012 06:38 AM - jayferd (Jay Adkisson)

<table>
<thead>
<tr>
<th>Status:</th>
<th>Rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority:</td>
<td>Normal</td>
</tr>
<tr>
<td>Assignee:</td>
<td>mame (Yusuke Endoh)</td>
</tr>
<tr>
<td>Target version:</td>
<td>2.6</td>
</tr>
<tr>
<td>ruby -v:</td>
<td>ruby 1.9.3p194 (2012-04-20 revision 35410) [x86_64-linux]</td>
</tr>
</tbody>
</table>

Description
To reproduce:

```ruby
>> require 'strscan'
>> ss = StringScanner.new("ab")
>> ss.scan(/./)
=> "a"
>> ss.scan(/^./) # expecting nil, since the head is in the middle of a line
=> "b"
```

As a workaround-that-didn't-work, I also tried looking behind for a newline, which seems never to match, even with the m flag.

```ruby
>> ss = StringScanner.new("a\nb")
>> ss.scan(/../m)
=> "a\n"
>> ss.scan(/(?<=\n)./m) # expecting "b", since the previous char is a newline
=> nil
```

History
#1 - 10/22/2012 06:07 AM - jayferd (Jay Adkisson)
Bump. Can someone confirm whether or not this is expected behavior?

#2 - 11/02/2012 08:33 PM - jayferd (Jay Adkisson)
I've now reproduced this in 2.0.0-preview1. Is anyone out there...?

#3 - 11/02/2012 10:11 PM - mame (Yusuke Endoh)
- Status changed from Open to Feedback
- Assignee set to mame (Yusuke Endoh)
- Target version set to 2.0.0

Sorry for late reply.

Honestly, it is difficult to say whether it is intended or a bug because there is no strscan maintainer. But since 1.8, StringScanner#scan has behaved so. I guess there is some existing programs that depends on the behavior.

Therefore, I'd like not to change the behavior unless there is no special reason. Do you have any special reason? Or do you just want to confirm?

--
Yusuke Endoh mame@tsg.ne.jp

#4 - 11/02/2012 10:12 PM - mame (Yusuke Endoh)
FYI: You may want to use StringScanner#beginning_of_line?.

--
Yusuke Endoh mame@tsg.ne.jp

#5 - 11/02/2012 11:30 PM - nobu (Nobuyoshi Nakada)
I guess it would be intentional. 
' and 'A' are 'beginning of the target' (and beginning of a line for \ anchors, so seems reasonable that it matches the beginning of the part to be matched.

#6 - 02/08/2013 11:55 PM - mame (Yusuke Endoh)
- Priority changed from Normal to 3
- Target version changed from 2.0.0 to 2.6

#7 - 05/31/2013 01:23 PM - mame (Yusuke Endoh)
- Status changed from Feedback to Rejected

I'm marking this ticket as closed due to no feedback.

--
Yusuke Endoh mame@tsg.ne.jp