

## Ruby master - Bug #5278

### REXML -- Malformed comment

09/06/2011 07:07 AM - tbf (Thomas Fritzsche)

<b>Status:</b> Closed	
<b>Priority:</b> Normal	
<b>Assignee:</b> kou (Kouhei Sutou)	
<b>Target version:</b> 1.9.3	
<b>ruby -v:</b> ruby 1.9.2p290 (2011-07-09 revision 32553) [x86_64-darwin11.1.0]	<b>Backport:</b>

#### Description

Hi Ruby-Team,

I use lib rexml for XML parsing. Kanjigidic2 XML-File: <http://www.csse.monash.edu.au/~jwb/kanjigidic2/> (I do not attach file because it is too large)

It works with version 1.8.7 but ParseException ("Malformed comment" is raised in lib/rexml/parsers/baseparser.rb

#### My Code looks like this:

```
require 'rexml/document'
require 'rexml/streamlistener'
class KanjiListener
include REXML::StreamListener
end
```

```
f = File.new("kanji.xml", "rb")
list = KanjiListener.new
```

#### REXML::Document.parse\_stream(f, list)

The used XML-File from above link has a comment section that looks like:

```
...
<!-- Version 1.6 - April 2008
This is the DTD of the XML-format kanji file combining information from
the KANJIDIC and KANJD212 files. It is intended to be largely self-
documenting, with each field being accompanied by an explanatory
comment.
-->
...
```

It's strange but the parser fails at "self- documented".

The issue comes up here (about line 345):

```
...
if md[0][2] == ?-
md = @source.match( COMMENT_PATTERN, true )

      case md[1]
      when /--/, /-$/
        raise REXML::ParseException.new("Malformed comment", @source)
      end
...

```

The MatchingData md[1] contains the complete comment and then regular expression /-\$/ matches.

From Debugging I guess the original Buffer is read by "readline" and somehow still includes the end-of-line markers.

I tried to open the original FileIO with different newline-parameters but nothing helped. I tried different ruby versions (incl. today's 1.9.3-head) but complete 1.9 seems to have the problem while 1.8 works.

I meanwhile converted to nokogiri XML-Parser and this works without problem on 1.9.x and I would expect that REXML could parse

this too. For test purpose I just changed a single character on this file so that "/-\$/" does not match "self-" in original XML file and than it works.

□□□□□□□□□□□□□□

### Associated revisions

#### Revision 0297247f - 02/06/2013 05:19 AM - usa (Usaku NAKAMURA)

merge revision(s) 33210,33212: [Backport #5278]

```
* lib/rexml/parsers/baseparser.rb, test/rexml/test_comment.rb:
  allow a single hyphen in comment. [Bug 5278]
  Reported by Thomas Fritzsche. Thanks!!!
  allow a single hyphen in comment. [Bug #5278] [ruby-core:39289]
```

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby\_1\_9\_3@39093 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

### History

#### #1 - 09/06/2011 10:37 AM - naruse (Yui NARUSE)

- Status changed from Open to Assigned
- Assignee set to kou (Kouhei Sutou)
- Target version set to 1.9.3

#### #2 - 09/07/2011 11:54 PM - kou (Kouhei Sutou)

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

Thanks for your report!  
I've fixed it in r33210.

#### #3 - 02/06/2013 08:53 PM - kou (Kouhei Sutou)

□□□□□

□□□□□□□□□□□□□□□□□□□□

<https://bugs.ruby-lang.org/issues/7764>

ChangeLog□□□□□□□□□□□□□□□□□□□□

<https://bugs.ruby-lang.org/projects/ruby-trunk/repository/revisions/33210>

□□□□□□□□□□□□□□□□□□□□

□lib/rexml/parsers/baseparser.rb□□□□□□□□□□□□

□□□□□□□□□□□□□□□□

In [20130206051927.E67E568693@sakura.atdot.net](mailto:20130206051927.E67E568693@sakura.atdot.net)

"[ruby-changes:27041] usa:r39093 (ruby\_1\_9\_3): merge revision(s) 33210,33212: [Backport #5278]" on Wed, 6 Feb 2013 14:19:27 +0900 (JST),  
usa [kou1@atdot.net](mailto:kou1@atdot.net) wrote:

usa 2013-02-06 14:19:18 +0900 (Wed, 06 Feb 2013)

New Revision: 39093

<http://svn.ruby-lang.org/cgi-bin/viewvc.cgi?view=rev&revision=39093>

Log:

merge revision(s) 33210,33212: [Backport #5278]

```
* lib/rexml/parsers/baseparser.rb, test/rexml/test_comment.rb:
  allow a single hyphen in comment. [Bug 5278]
  Reported by Thomas Fritzsche. Thanks!!!
  allow a single hyphen in comment. [Bug #5278] [ruby-core:39289]
```

Modified directories:

branches/ruby\_1\_9\_3/

Modified files:

branches/ruby\_1\_9\_3/ChangeLog

branches/ruby\_1\_9\_3/version.h

## Index: ruby\_1\_9\_3/ChangeLog

--- ruby\_1\_9\_3/ChangeLog (revision 39092)  
+++ ruby\_1\_9\_3/ChangeLog (revision 39093)  
@@ -1,3 +1,9 @@ [https://github.com/ruby/ruby/blob/trunk/ruby\\_1\\_9\\_3/ChangeLog#L1](https://github.com/ruby/ruby/blob/trunk/ruby_1_9_3/ChangeLog#L1)  
+Wed Feb 6 14:19:07 2013 Kouhei Sutou [kou@cozmixng.org](mailto:kou@cozmixng.org)  
+

- \* lib/rexml/parsers/baseparser.rb, test/rexml/test\_comment.rb:
- allow a single hyphen in comment. [Bug #5278] [ruby-core:39289]

• XXXXXXXXXXXXXXXXXXXX

+  
Wed Feb 6 14:14:38 2013 Nobuyoshi Nakada [nobu@ruby-lang.org](mailto:nobu@ruby-lang.org)

o file.c (realpath\_rec): prevent link from GC while link\_names refers Index: ruby\_1\_9\_3/version.h  
===== --- ruby\_1\_9\_3/version.h (revision 39092)  
+++ ruby\_1\_9\_3/version.h (revision 39093) @@ -1,5 +1,5 @@ [https://github.com/ruby/ruby/blob/trunk/ruby\\_1\\_9\\_3/version.h#L1](https://github.com/ruby/ruby/blob/trunk/ruby_1_9_3/version.h#L1)  
#define RUBY\_VERSION "1.9.3" -#define RUBY\_PATCHLEVEL 378 +#define RUBY\_PATCHLEVEL 379

#define RUBY\_RELEASE\_DATE "2013-02-06"  
#define RUBY\_RELEASE\_YEAR 2013

Property changes on: ruby\_1\_9\_3

Modified: svn:mergeinfo  
Merged /trunk:r33210,33212

--  
ML: [ruby-changes@quickml.atdot.net](mailto:ruby-changes@quickml.atdot.net)  
Info: <http://www.atdot.net/~ko1/quickml/>

#### #4 - 02/06/2013 09:53 PM - usa (Usaku NAKAMURA)

XXXXXXXXXXXX(X)XXXX

In message "[ruby-dev:46929] Re: [ruby-changes:27041] usa:r39093 (ruby\_1\_9\_3): merge revision(s) 33210,33212: [Backport #5278]" on Feb.06,2013 20:46:00, [kou@cozmixng.org](mailto:kou@cozmixng.org) wrote:

XXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXX  
<https://bugs.ruby-lang.org/issues/7764>

ChangeLogXXXXXXXXXXXXXXXXXXXXXXXXXXXX  
<https://bugs.ruby-lang.org/projects/ruby-trunk/repository/visions/33210>  
XXXXXXXXXXXXXXXXXXXXXXXXXXXX  
lib/rexml/parsers/baseparser.rbXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXX

XXXXXX

--  
U.Nakamura [usa@garbagecollect.jp](mailto:usa@garbagecollect.jp)

#### #5 - 02/06/2013 10:29 PM - kou (Kouhei Sutou)

XXXXXX

In [20130206132214.95E816EA62@zanzibar.garbagecollect.jp](mailto:20130206132214.95E816EA62@zanzibar.garbagecollect.jp)  
"[ruby-dev:46931] Re: [ruby-changes:27041] usa:r39093 (ruby\_1\_9\_3): merge revision(s) 33210,33212: [Backport #5278]" on Wed, 6 Feb 2013 22:22:14 +0900,  
"U.Nakamura" [usa@garbagecollect.jp](mailto:usa@garbagecollect.jp) wrote:

XXXXXXXXXXXXXXXXXXXXXXXXXXXX  
<https://bugs.ruby-lang.org/issues/7764>

ChangeLog  
<https://bugs.ruby-lang.org/projects/ruby-trunk/repository/revisions/33210>  
lib/rexml/parsers/baseparser.rb

svn merge