

Ruby master - Bug #3780

RDoc::Parser.binary? broken for some utf8 files longer than 1024 bytes

09/02/2010 06:21 AM - stepheneb (Stephen Bannasch)

Status: Closed	
Priority: Normal	
Assignee: drbrain (Eric Hodel)	
Target version: 1.9.3	
ruby -v: ruby 1.9.2p0 (2010-08-18 revision 29036) [x86_64-darwin10.4.0]	Backport:
Description	
<pre>=begin RDoc truncates files at 1024 bytes when checking if the file is binary. This will invalidate the file encoding if the file is truncated in the middle of a utf8 char and cause RDoc to exit. I found this problem when running rdoc on the ruby 1.9.2 source. \$ ruby -v ruby 1.9.2p0 (2010-08-18 revision 29036) [x86_64-darwin10.4.0] \$ rdoc --version rdoc 2.5.11 More description of the bug and a patch with a failing test is on this issue in RubyForge rdoc issue tracker. http://rubyforge.org/tracker/index.php?func=detail&aid=28525&group_id=627&atid=2472 The same issue appears to be in the 1_9 source, see: http://github.com/ruby/ruby/blob/trunk/lib/rdoc/parser.rb#L70 I find it confusing knowing where to create an RDoc issue: RubyForge or here -- so I've created an issue in both places. This gist: http://gist.github.com/561350 (possible_fix.rb) shows how I changed RDoc::Parser.binary? locally -- but I don't think it is correct to classify all utf8 files which are invalid when truncated at 1024 bytes as binary. That same gist (show_parsing_error.rb) also shows another strategy for solving the invalid encoding issue but there are probably better ways to determine if a file is binary. =end</pre>	

History

#1 - 09/02/2010 05:54 PM - tenderlovmaking (Aaron Patterson)

- Assignee set to drbrain (Eric Hodel)

```
=begin
```

```
=end
```

#2 - 09/08/2010 11:04 AM - drbrain (Eric Hodel)

```
=begin
```

RDoc 2.5.11 is newer than the version of RDoc than ships with Ruby 1.9.2.

RDoc 2.5.8 ships with Ruby 1.9.2.

Can you confirm that this bug exists in the default RDoc that ships with 1.9.2?

```
=end
```

#3 - 09/08/2010 07:15 PM - stepheneb (Stephen Bannasch)

```
=begin
```

Interesting ... the problem does not occur when running rdoc included in ruby built with the v1_9_2_0 tag. I had thought it would -- but the RDoc::Parser.binary? method I reference above which I believe causes the problem: <http://github.com/ruby/ruby/blob/trunk/lib/rdoc/parser.rb#L70> is from trunk -- appears to be identical??

```
$ ./bin/ruby --version
```

ruby 1.9.2p0 (2010-08-18 revision 29036) [x86_64-darwin10.4.0]

```
$. /bin/rdoc --version
rdoc 2.5.8
```

```
$ rm -rf ~/Desktop/rdoc; ./bin/rdoc -o ~/Desktop/rdoc ~/.rvm/rubies/ruby-1.9.2-p0/lib/ruby/1.9.1/
Parsing sources...
/Users/stephen/.rvm/rubies/ruby-1.9.2-p0/lib/ruby/1.9.1/irb/inspector.rb:36:36: Couldn't find INSPECTORS. Assuming it's a module
/Users/stephen/.rvm/rubies/ruby-1.9.2-p0/lib/ruby/1.9.1/singleton.rb:238:11: Couldn't find Yup. Assuming it's a module
/Users/stephen/.rvm/rubies/ruby-1.9.2-p0/lib/ruby/1.9.1/tk/font.rb:41:27: Couldn't find SYSTEM_FONT_NAMES. Assuming it's a module
/Users/stephen/.rvm/rubies/ruby-1.9.2-p0/lib/ruby/1.9.1/tk.rb:67:30: Couldn't find Tk_CMDTBL. Assuming it's a module
/Users/stephen/.rvm/rubies/ruby-1.9.2-p0/lib/ruby/1.9.1/tk.rb:72:31: Couldn't find Tk_WINDOWS. Assuming it's a module
100% [877/877] /Users/stephen/.rvm/rubies/ruby-1.9.2-p0/lib/ruby/1.9.1/yaml.rb
```

Generating Darkfish...

```
Files: 877
Classes: 1647 ( 1138 undocumented)
Constants: 1894 ( 1630 undocumented)
Modules: 444 ( 314 undocumented)
Methods: 12982 ( 9305 undocumented)
26.99% documented
```

Elapsed: 285.4s

=end

#4 - 09/09/2010 02:36 AM - drbrain (Eric Hodel)

- Category set to lib
- Status changed from Open to Assigned
- Priority changed from Normal to 3
- Target version set to 1.9.3

=begin

Ok, thanks for the confirmation of where the problem occurs.

I've been adding proper encoding support to RDoc and it reveals that the current implementation is naive.

The next release should work properly on 1.9.

=end

#5 - 02/03/2011 10:42 AM - drbrain (Eric Hodel)

- Status changed from Assigned to Closed

=begin

Fixed by import of RDoc 3.5

=end