

Ruby - Feature #20210

Invalid source encoding raises ArgumentError, not SyntaxError

01/25/2024 02:28 AM - kddnewton (Kevin Newton)

Status:	Closed	
Priority:	Normal	
Assignee:		
Target version:		
<p>Description</p> <p>I was hoping we could change the error that is raised when an invalid source encoding is found from an ArgumentError to a SyntaxError.</p> <p>First let me say, if this isn't possible for backward compatibility, I understand. Please do not take this as me not caring about backward compatibility.</p> <p>Right now, if you have the script # encoding: foo\n"bar", it will raise an ArgumentError, not a SyntaxError. If there are other syntax errors in the file, there's no way to concat them together to give feedback to the user. If a user wants to consistently handle the errors coming back from a parse, they currently have to rescue ArgumentError and SyntaxError.</p> <p>Ideally it would all be SyntaxError, so we could handle them consistently and append all errors together.</p>		

History

#1 - 01/25/2024 06:54 AM - nobu (Nobuyoshi Nakada)

- Tracker changed from Misc to Feature

I don't remember the reason to select ArgumentError, SyntaxError feels more reasonable.
<https://github.com/ruby/ruby/pull/9701>

#2 - 01/29/2024 07:57 AM - yui-ken (Kaneko Yuichiro)

I'm wondering which encoding should be used if the parser hits invalid source encoding like # coding: foo. I think it's needed to clarify which encoding is assumed on this ticket.

#3 - 02/01/2024 12:13 AM - naruse (Yui NARUSE)

Parsing entire source code with wrong encoding is not reasonable because in some encoding including SJIS (Windows-31J) parsing result won't be valid because some multibyte character may include ASCII character in the trailing byte in the encoding. A developer need to fix the encoding first.

#4 - 02/01/2024 01:52 AM - Anonymous

Hi. One question:

When parsing begins, what encoding do Prism and parse.y use by default?

#5 - 02/01/2024 02:23 AM - kddnewton (Kevin Newton)

[@naruse \(Yui NARUSE\)](#) I'm fine exiting immediately, I was just hoping to make it a syntax error.

[@Edwing123](#) By default Ruby source assumes UTF-8 unless told otherwise by a magic comment or a command line option.

#6 - 02/14/2024 11:20 AM - mame (Yusuke Endoh)

Discussed at the dev meeting.

We need a good reason to introduce incompatibility. You say you are fine with the current behavior (exiting immediately), Then, we can't see no reason to change it.

This is just my idea: it would be great for prism, as a library, to continue parsing even with an invalid encoding magic comment (maybe as ASCII-8BIT?), but it would be good to keep the behavior of Ruby interpreter as possible.

#7 - 02/14/2024 01:36 PM - kddnewton (Kevin Newton)

- Status changed from Open to Closed

I think that makes sense! Let's keep it as an argument error. Prism will keep parsing for now, but raise the right error.