

Ruby trunk - Bug #8342

IO.readlines ignores Encoding.default_internal if Encoding.default_external is ASCII-8BIT

04/28/2013 09:13 PM - leocassarani (Leo Cassarani)

Status:	Closed	
Priority:	Normal	
Assignee:	naruse (Yui NARUSE)	
Target version:	2.1.0	
ruby -v:	1.9.3	Backport: 1.9.3: UNKNOWN, 2.0.0: UNKNOWN
Description		
<p>Under normal circumstances, IO.readlines will transcode from Encoding.default_external to Encoding.default_internal:</p> <pre>File.open('hi', 'w') { f f.puts "hello\n" } Encoding.default_external = Encoding::US_ASCII Encoding.default_internal = Encoding::UTF_8 puts IO.readlines('hi').first.encoding #=> UTF-8</pre> <p>However, when Encoding.default_external is set to ASCII-8BIT, IO.readlines will always use ASCII-8BIT, regardless of what Encoding.default_internal is set to:</p> <pre>File.open('hi', 'w') { f f.puts "hello\n" } Encoding.default_external = Encoding::ASCII_8BIT Encoding.default_internal = Encoding::UTF_8 puts IO.readlines('hi').first.encoding #=> ASCII-8BIT</pre> <p>Using IO#gets instead of IO.readlines will produce the same behaviour.</p>		

Associated revisions

Revision 4834aa53 - 05/08/2013 08:47 AM - naruse (Yui NARUSE)

- io.c (rb_io_ext_int_to_encs): ignore internal encoding if external encoding is ASCII-8BIT. [Bug #8342]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@40610 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 40610 - 05/08/2013 08:47 AM - naruse (Yui NARUSE)

- io.c (rb_io_ext_int_to_encs): ignore internal encoding if external encoding is ASCII-8BIT. [Bug #8342]

Revision 40610 - 05/08/2013 08:47 AM - naruse (Yui NARUSE)

- io.c (rb_io_ext_int_to_encs): ignore internal encoding if external encoding is ASCII-8BIT. [Bug #8342]

Revision 40610 - 05/08/2013 08:47 AM - naruse (Yui NARUSE)

- io.c (rb_io_ext_int_to_encs): ignore internal encoding if external encoding is ASCII-8BIT. [Bug #8342]

Revision 40610 - 05/08/2013 08:47 AM - naruse (Yui NARUSE)

- io.c (rb_io_ext_int_to_encs): ignore internal encoding if external encoding is ASCII-8BIT. [Bug #8342]

Revision 40610 - 05/08/2013 08:47 AM - naruse (Yui NARUSE)

- io.c (rb_io_ext_int_to_encs): ignore internal encoding if external encoding is ASCII-8BIT. [Bug #8342]

Revision 40610 - 05/08/2013 08:47 AM - naruse (Yui NARUSE)

- io.c (rb_io_ext_int_to_encs): ignore internal encoding if external encoding is ASCII-8BIT. [Bug #8342]

History

#1 - 04/29/2013 09:35 AM - nobu (Nobuyoshi Nakada)

- Category set to M17N
- Status changed from Open to Assigned
- Assignee set to naruse (Yui NARUSE)
- Target version set to 2.1.0

Seems intended behavior to me.

#2 - 04/29/2013 02:00 PM - naruse (Yui NARUSE)

- Status changed from Assigned to Rejected

If external encoding is ASCII-8BIT, the input content is considered as binary.
It is out of text encoding conversion and its encoding kept as ASCII-8BIT even if default_internal is set.

#3 - 05/02/2013 11:12 PM - leocassarani (Leo Cassarani)

Thanks naruse. However, this seems inconsistent with the way encodings are handled for individual IO instances. For example:

```
io = File.open('hi', :encoding => "ascii-8bit:utf-16")
puts io.gets.encoding
```

=> UTF-16

This happens even if Encoding.default_external is set to ASCII-8BIT before opening the file.

#4 - 05/02/2013 11:56 PM - naruse (Yui NARUSE)

- Status changed from Rejected to Assigned

leocassarani (Leo Cassarani) wrote:

Thanks naruse. However, this seems inconsistent with the way encodings are handled for individual IO instances. For example:

```
io = File.open('hi', :encoding => "ascii-8bit:utf-16")
puts io.gets.encoding
```

=> UTF-16

This happens even if Encoding.default_external is set to ASCII-8BIT before opening the file.

That side sounds buggy

#5 - 05/08/2013 05:47 PM - naruse (Yui NARUSE)

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

This issue was solved with changeset [r40610](#).
Leo, thank you for reporting this issue.
Your contribution to Ruby is greatly appreciated.
May Ruby be with you.

-
- io.c (rb_io_ext_int_to_encs): ignore internal encoding if external encoding is ASCII-8BIT. [Bug [#8342](#)]