

Ruby master - Bug #13292

Invalid encodings in UTF-32

03/08/2017 01:14 PM - rbjl (Jan Lejis)

Status: Closed	
Priority: Normal	
Assignee:	
Target version:	
ruby -v: ruby 2.4.0p0 (2016-12-24 revision 57164) [x86_64-linux]	Backport: 2.2: DONE, 2.3: DONE, 2.4: DONE
Description Ruby is very strict about valid UTF-8 encodings, which is great. Strings that encode surrogates or too large codepoints are not valid. However, in UTF-32, it is possible to encode such values, and Ruby treats them as valid: Example 1 (too large value) <pre>a = [0, 0, 17, 0].pack("C*").force_encoding("UTF-32LE") #=> "\u{110000}" a.valid_encoding? # => true</pre> Example 2 (surrogate) <pre>b = [0, 216, 0, 0].pack("C*").force_encoding("UTF-32LE") # => "\uD800" b.valid_encoding? #=> true</pre> The behaviour should be changed to String#valid_encoding? reporting false For reference: http://unicode.org/versions/Unicode9.0.0/UnicodeStandard-9.0.pdf (page 71)	

Associated revisions

Revision 4171ed6c - 03/09/2017 02:04 AM - nobu (Nobuyoshi Nakada)

fix UTF-32 valid_encoding?

- enc/utf_32be.c (utf32be_mbc_enc_len): check arguments precisely. [ruby-core:79966] [Bug #13292]
- enc/utf_32le.c (utf32le_mbc_enc_len): ditto.
- regenc.h (UNICODE_VALID_CODEPOINT_P): predicate for valid Unicode codepoints.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@57816 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 57816 - 03/09/2017 02:04 AM - nobu (Nobuyoshi Nakada)

fix UTF-32 valid_encoding?

- enc/utf_32be.c (utf32be_mbc_enc_len): check arguments precisely. [ruby-core:79966] [Bug #13292]
- enc/utf_32le.c (utf32le_mbc_enc_len): ditto.
- regenc.h (UNICODE_VALID_CODEPOINT_P): predicate for valid Unicode codepoints.

Revision 57816 - 03/09/2017 02:04 AM - nobu (Nobuyoshi Nakada)

fix UTF-32 valid_encoding?

- `enc/utf_32be.c (utf32be_mbc_enc_len)`: check arguments precisely. [ruby-core:79966] [Bug #13292]
- `enc/utf_32le.c (utf32le_mbc_enc_len)`: ditto.
- `regenc.h (UNICODE_VALID_CODEPOINT_P)`: predicate for valid Unicode codepoints.

Revision 57816 - 03/09/2017 02:04 AM - nobu (Nobuyoshi Nakada)

fix UTF-32 `valid_encoding?`

- `enc/utf_32be.c (utf32be_mbc_enc_len)`: check arguments precisely. [ruby-core:79966] [Bug #13292]
- `enc/utf_32le.c (utf32le_mbc_enc_len)`: ditto.
- `regenc.h (UNICODE_VALID_CODEPOINT_P)`: predicate for valid Unicode codepoints.

Revision 35fde4da - 03/09/2017 02:10 AM - nobu (Nobuyoshi Nakada)

fix UTF-32 `valid_encoding?`

- `test/ruby/test_io_m17n.rb (TestIO_M17N#test_puts_widechar)`: do not use invalid codepoint. [ruby-core:79966] [Bug #13292]

git-svn-id: [svn+ssh://ci.ruby-lang.org/ruby/trunk@57817](https://ci.ruby-lang.org/ruby/trunk@57817) b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 57817 - 03/09/2017 02:10 AM - nobu (Nobuyoshi Nakada)

fix UTF-32 `valid_encoding?`

- `test/ruby/test_io_m17n.rb (TestIO_M17N#test_puts_widechar)`: do not use invalid codepoint. [ruby-core:79966] [Bug #13292]

Revision 57817 - 03/09/2017 02:10 AM - nobu (Nobuyoshi Nakada)

fix UTF-32 `valid_encoding?`

- `test/ruby/test_io_m17n.rb (TestIO_M17N#test_puts_widechar)`: do not use invalid codepoint. [ruby-core:79966] [Bug #13292]

Revision 57817 - 03/09/2017 02:10 AM - nobu (Nobuyoshi Nakada)

fix UTF-32 `valid_encoding?`

- `test/ruby/test_io_m17n.rb (TestIO_M17N#test_puts_widechar)`: do not use invalid codepoint. [ruby-core:79966] [Bug #13292]

Revision acfebb41 - 03/12/2017 05:22 PM - naruse (Yui NARUSE)

merge revision(s) 57816,57817: [Backport #13292]

```
fix UTF-32 valid_encoding?
```

```
* enc/utf_32be.c (utf32be_mbc_enc_len): check arguments precisely.
[ruby-core:79966] [Bug #13292]
```

```
* enc/utf_32le.c (utf32le_mbc_enc_len): ditto.
```

```
* regenc.h (UNICODE_VALID_CODEPOINT_P): predicate for valid
Unicode codepoints.
fix UTF-32 valid_encoding?
```

```
* test/ruby/test_io_m17n.rb (TestIO_M17N#test_puts_widechar): do
not use invalid codepoint. [ruby-core:79966] [Bug #13292]
```

git-svn-id: [svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_2_4@57935](https://ci.ruby-lang.org/ruby/branches/ruby_2_4@57935) b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 57935 - 03/12/2017 05:22 PM - naruse (Yui NARUSE)

merge revision(s) 57816,57817: [Backport #13292]

```
fix UTF-32 valid_encoding?
```

```
* enc/utf_32be.c (utf32be_mbc_enc_len): check arguments precisely.
[ruby-core:79966] [Bug #13292]

* enc/utf_32le.c (utf32le_mbc_enc_len): ditto.

* regenc.h (UNICODE_VALID_CODEPOINT_P): predicate for valid
Unicode codepoints.
fix UTF-32 valid_encoding?

* test/ruby/test_io_m17n.rb (TestIO_M17N#test_puts_widechar): do
not use invalid codepoint. [ruby-core:79966] [Bug #13292]
```

Revision ad075f69 - 03/25/2017 05:01 PM - usa (Usaku NAKAMURA)

merge revision(s) 57816,57817: [Backport #13292]

```
fix UTF-32 valid_encoding?

* enc/utf_32be.c (utf32be_mbc_enc_len): check arguments precisely.
[ruby-core:79966] [Bug #13292]

* enc/utf_32le.c (utf32le_mbc_enc_len): ditto.

* regenc.h (UNICODE_VALID_CODEPOINT_P): predicate for valid
Unicode codepoints.
fix UTF-32 valid_encoding?

* test/ruby/test_io_m17n.rb (TestIO_M17N#test_puts_widechar): do
not use invalid codepoint. [ruby-core:79966] [Bug #13292]
```

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_2_2@58103 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 58103 - 03/25/2017 05:01 PM - usa (Usaku NAKAMURA)

merge revision(s) 57816,57817: [Backport #13292]

```
fix UTF-32 valid_encoding?

* enc/utf_32be.c (utf32be_mbc_enc_len): check arguments precisely.
[ruby-core:79966] [Bug #13292]

* enc/utf_32le.c (utf32le_mbc_enc_len): ditto.

* regenc.h (UNICODE_VALID_CODEPOINT_P): predicate for valid
Unicode codepoints.
fix UTF-32 valid_encoding?

* test/ruby/test_io_m17n.rb (TestIO_M17N#test_puts_widechar): do
not use invalid codepoint. [ruby-core:79966] [Bug #13292]
```

Revision 909331e2 - 03/27/2017 05:34 PM - nagachika (Tomoyuki Chikanaga)

merge revision(s) 57816,57817: [Backport #13292]

```
fix UTF-32 valid_encoding?

* enc/utf_32be.c (utf32be_mbc_enc_len): check arguments precisely.
[ruby-core:79966] [Bug #13292]

* enc/utf_32le.c (utf32le_mbc_enc_len): ditto.

* regenc.h (UNICODE_VALID_CODEPOINT_P): predicate for valid
Unicode codepoints.
fix UTF-32 valid_encoding?

* test/ruby/test_io_m17n.rb (TestIO_M17N#test_puts_widechar): do
not use invalid codepoint. [ruby-core:79966] [Bug #13292]
```

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_2_3@58183 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 58183 - 03/27/2017 05:34 PM - nagachika (Tomoyuki Chikanaga)

merge revision(s) 57816,57817: [Backport #13292]

```
fix UTF-32 valid_encoding?

* enc/utf_32be.c (utf32be_mbc_enc_len): check arguments precisely.
[ruby-core:79966] [Bug #13292]

* enc/utf_32le.c (utf32le_mbc_enc_len): ditto.

* regenc.h (UNICODE_VALID_CODEPOINT_P): predicate for valid
Unicode codepoints.
fix UTF-32 valid_encoding?

* test/ruby/test_io_m17n.rb (TestIO_M17N#test_puts_widechar): do
not use invalid codepoint. [ruby-core:79966] [Bug #13292]
```

History

#1 - 03/09/2017 02:04 AM - nobu (Nobuyoshi Nakada)

- Status changed from Open to Closed

Applied in changeset r57816.

fix UTF-32 valid_encoding?

- enc/utf_32be.c (utf32be_mbc_enc_len): check arguments precisely.
[ruby-core:79966] [Bug #13292]
- enc/utf_32le.c (utf32le_mbc_enc_len): ditto.
- regenc.h (UNICODE_VALID_CODEPOINT_P): predicate for valid
Unicode codepoints.

#2 - 03/12/2017 05:22 PM - naruse (Yui NARUSE)

- Backport changed from 2.2: UNKNOWN, 2.3: UNKNOWN, 2.4: UNKNOWN to 2.2: UNKNOWN, 2.3: UNKNOWN, 2.4: DONE

ruby_2_4 r57935 merged revision(s) 57816,57817.

#3 - 03/14/2017 01:06 AM - nagachika (Tomoyuki Chikanaga)

- Backport changed from 2.2: UNKNOWN, 2.3: UNKNOWN, 2.4: DONE to 2.2: REQUIRED, 2.3: REQUIRED, 2.4: DONE

#4 - 03/25/2017 05:01 PM - usa (Usaku NAKAMURA)

- Backport changed from 2.2: REQUIRED, 2.3: REQUIRED, 2.4: DONE to 2.2: DONE, 2.3: REQUIRED, 2.4: DONE

ruby_2_2 r58103 merged revision(s) 57816,57817.

#5 - 03/27/2017 05:34 PM - nagachika (Tomoyuki Chikanaga)

- Backport changed from 2.2: DONE, 2.3: REQUIRED, 2.4: DONE to 2.2: DONE, 2.3: DONE, 2.4: DONE

ruby_2_3 r58183 merged revision(s) 57816,57817.