Ruby master - Bug #6799
Digest::*:hexdigest returns an ASCII-8BIT String
07/27/2012 06:17 AM - Eregon (Benoit Daloze)

Status: Closed
Priority: Normal
Assignee: Eregon (Benoit Daloze)
Target version: 2.0.0
ruby -v:
ruby 2.0.0dev (2012-07-15 trunk 36395)
[x86_64-darwin10.8.0]

Description
$ ruby -rdigest/sha1 -e 'p Digest::SHA1.hexdigest("a").encoding'

#Encoding:ASCII-8BIT

I'm happy to provide a patch.

Associated revisions
Revision 61b624d1 - 08/01/2012 01:30 PM - Eregon (Benoit Daloze)
ext/digest/digest.c (hexencode_str_new): return an ASCII string

- test/digest: tests for all kind of digests encodings [ruby-core:46792][Bug #6799]
git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@36588 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 36588 - 08/01/2012 01:30 PM - Eregon (Benoit Daloze)
ext/digest/digest.c (hexencode_str_new): return an ASCII string

- test/digest: tests for all kind of digests encodings [ruby-core:46792][Bug #6799]

Revision 36588 - 08/01/2012 01:30 PM - Eregon (Benoit Daloze)
ext/digest/digest.c (hexencode_str_new): return an ASCII string

- test/digest: tests for all kind of digests encodings [ruby-core:46792][Bug #6799]

Revision 36588 - 08/01/2012 01:30 PM - Eregon (Benoit Daloze)
ext/digest/digest.c (hexencode_str_new): return an ASCII string

- test/digest: tests for all kind of digests encodings [ruby-core:46792][Bug #6799]

Revision 36588 - 08/01/2012 01:30 PM - Eregon (Benoit Daloze)
ext/digest/digest.c (hexencode_str_new): return an ASCII string

- test/digest: tests for all kind of digests encodings [ruby-core:46792][Bug #6799]

Revision 36588 - 08/01/2012 01:30 PM - Eregon (Benoit Daloze)
ext/digest/digest.c (hexencode_str_new): return an ASCII string

- test/digest: tests for all kind of digests encodings [ruby-core:46792][Bug #6799]

Revision 36588 - 08/01/2012 01:30 PM - Eregon (Benoit Daloze)
ext/digest/digest.c (hexencode_str_new): return an ASCII string

- test/digest: tests for all kind of digests encodings [ruby-core:46792][Bug #6799]

Revision 36588 - 08/01/2012 01:30 PM - Eregon (Benoit Daloze)
ext/digest/digest.c (hexencode_str_new): return an ASCII string

- test/digest: tests for all kind of digests encodings [ruby-core:46792][Bug #6799]

Revision 36588 - 08/01/2012 01:30 PM - Eregon (Benoit Daloze)
ext/digest/digest.c (hexencode_str_new): return an ASCII string

- test/digest: tests for all kind of digests encodings [ruby-core:46792][Bug #6799]

Revision b0e40509 - 10/19/2012 01:13 PM - Eregon (Benoit Daloze)
set encoding to ASCII for appropriate String#unpack modifiers

- pack.c (pack_unpack): set encoding of the ‘H’,’h’,’B’ and ‘B’ modifiers to US-ASCII.
set encoding to ASCII for appropriate String#unpack modifiers

pack.c (pack_unpack): set encoding of the 'H','h','B' and 'b' modifiers to US-ASCII.

test/ruby/test_pack.rb: tests for the above. [ruby-core:47653][Bug #7050]
test/test_securerandom.rb: tests for SecureRandom.hex from tenderlove. [ruby-core:46792][Bug #6799]

Revision 37269 - 10/19/2012 01:13 PM - Eregon (Benoit Daloze)

set encoding to ASCII for appropriate String#unpack modifiers

pack.c (pack_unpack): set encoding of the 'H','h','B' and 'b' modifiers to US-ASCII.

test/ruby/test_pack.rb: tests for the above. [ruby-core:47653][Bug #7050]
test/test_securerandom.rb: tests for SecureRandom.hex from tenderlove. [ruby-core:46792][Bug #6799]

Revision 37269 - 10/19/2012 01:13 PM - Eregon (Benoit Daloze)

set encoding to ASCII for appropriate String#unpack modifiers

pack.c (pack_unpack): set encoding of the 'H','h','B' and 'b' modifiers to US-ASCII.

test/ruby/test_pack.rb: tests for the above. [ruby-core:47653][Bug #7050]
test/test_securerandom.rb: tests for SecureRandom.hex from tenderlove. [ruby-core:46792][Bug #6799]

Revision 37269 - 10/19/2012 01:13 PM - Eregon (Benoit Daloze)

set encoding to ASCII for appropriate String#unpack modifiers

pack.c (pack_unpack): set encoding of the 'H','h','B' and 'b' modifiers to US-ASCII.

test/ruby/test_pack.rb: tests for the above. [ruby-core:47653][Bug #7050]
test/test_securerandom.rb: tests for SecureRandom.hex from tenderlove. [ruby-core:46792][Bug #6799]

Revision 37269 - 10/19/2012 01:13 PM - Eregon (Benoit Daloze)

set encoding to ASCII for appropriate String#unpack modifiers

pack.c (pack_unpack): set encoding of the 'H','h','B' and 'b' modifiers to US-ASCII.

test/ruby/test_pack.rb: tests for the above. [ruby-core:47653][Bug #7050]
test/test_securerandom.rb: tests for SecureRandom.hex from tenderlove. [ruby-core:46792][Bug #6799]

Revision 37269 - 10/19/2012 01:13 PM - Eregon (Benoit Daloze)

set encoding to ASCII for appropriate String#unpack modifiers

pack.c (pack_unpack): set encoding of the 'H','h','B' and 'b' modifiers to US-ASCII.

test/ruby/test_pack.rb: tests for the above. [ruby-core:47653][Bug #7050]
test/test_securerandom.rb: tests for SecureRandom.hex from tenderlove. [ruby-core:46792][Bug #6799]

Revision 37269 - 10/19/2012 01:13 PM - Eregon (Benoit Daloze)

set encoding to ASCII for appropriate String#unpack modifiers

pack.c (pack_unpack): set encoding of the 'H','h','B' and 'b' modifiers to US-ASCII.

test/ruby/test_pack.rb: tests for the above. [ruby-core:47653][Bug #7050]
test/test_securerandom.rb: tests for SecureRandom.hex from tenderlove. [ruby-core:46792][Bug #6799]

Revision 37269 - 10/19/2012 01:13 PM - Eregon (Benoit Daloze)

set encoding to ASCII for appropriate String#unpack modifiers

pack.c (pack_unpack): set encoding of the 'H','h','B' and 'b' modifiers to US-ASCII.

test/ruby/test_pack.rb: tests for the above. [ruby-core:47653][Bug #7050]
test/test_securerandom.rb: tests for SecureRandom.hex from tenderlove. [ruby-core:46792][Bug #6799]

History

#1 - 07/27/2012 10:21 AM - duerst (Martin Dürrst)
So what should it return? If you create a bug, saying what the expected behavior is (and also why you expect it) is quite important.

#2 - 07/27/2012 11:08 AM - shyouhei (Shyouhei Urabe)
- Status changed from Open to Feedback

#3 - 07/27/2012 07:34 PM - Eregon (Benoit Daloze)
duerst (Martin Dürrst) wrote:

So what should it return? If you create a bug, saying what the expected behavior is (and also why you expect it) is quite important.

I thought it was clear enough, but indeed I should have mentioned it.
An hexadecimal String like this should have the US-ASCII encoding, because, by definition, it only has ASCII characters ([a-f0-9]).

This for example, makes YAML dump it as a binary String, and may cause other problems as the String is believed to be "binary", while it's not (#digest on the other hand has the ASCII-8BIT encoding, which is right).
On Fri, Jul 27, 2012 at 07:34:30PM +0900, Eregon (Benoit Daloze) wrote:

Issue #6799 has been updated by Eregon (Benoit Daloze).

duerst (Martin Dürst) wrote:

So what should it return? [If you create a bug, saying what the expected behavior is (and also why you expect it) is quite important.]

I thought it was clear enough, but indeed I should have mentioned it.
An hexadecimal String like this should have the US-ASCII encoding, because, by definition, it only has ASCII characters ([a-f0-9]).

FWIW, I would like this feature too. In Rails, we have to work around this issue in some database drivers because the ASCII-8BIT string will be considered to be binary.

--
Aaron Patterson
http://tenderlovemaking.com/

Given that the string has only ASCII characters, an encoding of US-ASCII indeed seems best. Benoit, unless somebody objects soon, I think you should go ahead with your patch.

This issue was solved with changeset r36588.
Benoit, thank you for reporting this issue.
Your contribution to Ruby is greatly appreciated.
May Ruby be with you.

ext/digest/digest.c (hexencode_str_new): return an ASCII string

- test/digest: tests for all kind of digests encodings [ruby-core:46792][Bug #6799]

Given that the string has only ASCII characters, an encoding of US-ASCII indeed seems best. Benoit, unless somebody objects soon, I think you should go ahead with your patch.

Thank you, I just did.

Could we do this for SecureRandom.hex as well? I've attached a patch. /cc akr

#10 - 09/14/2012 07:35 PM - Eregon (Benoit Daloze)
Aaron (Aaron Kelly): Maybe String#unpack should be fixed instead to return an ASCII string for the 'H', 'h', 'B' and 'b' directives?
It makes sense to me since these will always contain 0-9a-f (0-1 for 'B','b') chars.
This is already done for the 'm', 'M' and 'u' directives in Array#pack.
What do you think?

P.S.: Does anyone know the rationale behind which method to use for 'u','m','M' and 'H','h','B','b'?
To have a base64 representation you need [s].pack('m') (and so #pack is used for 'u','m','M' to obtain the specified format), but for 'H','h','B','b' you need s.unpack(directive). I think the latter makes more sense when compared to other types such as 'l', in which #unpack is used to obtain the desired format.

#11 - 09/15/2012 03:53 PM - Anonymous

- File noname added

On Fri, Sep 14, 2012 at 07:35:46PM +0900, Eregon (Benoit Daloze) wrote:

Issue #6799 has been updated by Eregon (Benoit Daloze).

Aaron (Aaron Kelly): Maybe String#unpack should be fixed instead to return an ASCII string for the 'H', 'h', 'B' and 'b' directives?
It makes sense to me since these will always contain 0-9a-f (0-1 for 'B','b') chars.
This is already done for the 'm', 'M' and 'u' directives in Array#pack.
What do you think?

I agree (and would prefer that). It just seems like a larger change than making this method work for me. :-)

--
Aaron Patterson
http://tenderlovemaking.com/

Files

<table>
<thead>
<tr>
<th>File Name</th>
<th>Size</th>
<th>Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001-ext-digest-digest.c-hexencode_str_new-return-an-ASCI.patch</td>
<td>2.59 KB</td>
<td>07/27/2012</td>
<td>Eregon (Benoit Daloze)</td>
</tr>
<tr>
<td>noname</td>
<td>500 Bytes</td>
<td>07/29/2012</td>
<td>Anonymous</td>
</tr>
<tr>
<td>secure_random.patch</td>
<td>863 Bytes</td>
<td>09/06/2012</td>
<td>tenderlovemaking (Aaron Patterson)</td>
</tr>
<tr>
<td>noname</td>
<td>500 Bytes</td>
<td>09/15/2012</td>
<td>Anonymous</td>
</tr>
</tbody>
</table>