

Ruby - Bug #10300

Encoding error in conversion from UTF-16LE to UTF-8 to CP850

09/27/2014 03:01 PM - ggrossetie (Guillaume GROSSETIE)

Status:	Closed	
Priority:	Normal	
Assignee:	windows	
Target version:	2.2.0	
ruby -v:	2.1.3p242 (2014-09-19 revision 47630) [x64-mingw32]	Backport: 2.0.0: DONTNEED, 2.1: REQUIRED
Description		
Hello,		
I downloaded Ruby 2.1.3 from http://rubyinstaller.org/downloads/ and tried to install gems:		
<pre>\$ gem install asciidoctor ERROR: While executing gem ... (Encoding::UndefinedConversionError) U+2019 to CP850 in conversion from UTF-16LE to UTF-8 to CP850</pre>		
I googled the error and found a number of "solutions":		
<pre>\$ gem install asciidoctor -E utf-8 --no-rdoc \$ LC_ALL=fr.FR.UTF-8 LANG= gem install asciidoctor \$ export LC_CTYPE=utf-8 \$ export RUBYOPT='-E utf-8' \$ ruby -e 'p Encoding.default_external' #<Encoding:UTF-8></pre>		
The Encoding.default_external was now on UTF-8 but the error persisted.		
My environment:		
<pre>\$ gem env RubyGems Environment: - RUBYGEMS VERSION: 2.2.2 - RUBY VERSION: 2.1.3 (2014-09-19 patchlevel 242) [x64-mingw32] - INSTALLATION DIRECTORY: c:/Ruby21-x64/lib/ruby/gems/2.1.0 - RUBY EXECUTABLE: c:/Ruby21-x64/bin/ruby.exe - EXECUTABLE DIRECTORY: c:/Ruby21-x64/bin - SPEC CACHE DIRECTORY: c:/Users/gg1504en/.gem/specs - RUBYGEMS PLATFORMS: - ruby - x64-mingw32 - GEM PATHS: - c:/Ruby21-x64/lib/ruby/gems/2.1.0 - c:/Users/gg1504en/.gem/ruby/2.1.0 - GEM CONFIGURATION: - :update_sources => true - :verbose => true - :backtrace => false - :bulk_threshold => 1000 - REMOTE SOURCES: - https://rubygems.org/ - SHELL PATH: - c:\Users\gg1504en\bin - . - C:\dev\softs\git\local\bin - C:\dev\softs\git\mingw\bin - C:\dev\softs\git\bin - c:\progra~1\oracle\ora_10.2.0_clt\bin - c:\Windows\system32 - c:\Windows - c:\Windows\System32\Wbem</pre>		

```

- c:\Windows\System32\WindowsPowerShell\v1.0\
- c:\Program Files (x86)\QuickTime Alternative\QTSystem
- c:\Program Files (x86)\Microsoft Application Virtualization Client
- c:\Windows\system32\BioRTime
- c:\Windows\SysWOW64\BioRTime
- c:\dev\softs\java\jdk1.7.0_55\bin
- c:\dev\softs\maven-3.0.5\bin
- c:\Program Files (x86)\GNU\GnuPG\pub
- c:\Ruby21-x64\bin

```

```

$ ruby -v
ruby 2.1.3p242 (2014-09-19 revision 47630) [x64-mingw32]
$ gem -v
2.2.2

```

I turned on trace:

```

$ gem install --backtrace -V --no-ri --no-rdoc asciidoctor
ERROR: While executing gem ... (Encoding::UndefinedConversionError)
      U+2019 to CP850 in conversion from UTF-16LE to UTF-8 to CP850
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/registry.rb:178:in `encode'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/registry.rb:178:in `initialize'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/registry.rb:238:in `exception'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/registry.rb:238:in `raise'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/registry.rb:238:in `check'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/registry.rb:300:in `EnumKey'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/registry.rb:594:in `each_key'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/resolv.rb:85:in `block (2 levels) in get_info'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/registry.rb:422:in `open'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/registry.rb:529:in `open'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/resolv.rb:84:in `block in get_info'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/registry.rb:422:in `open'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/registry.rb:529:in `open'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/resolv.rb:61:in `get_info'
      c:/Ruby21-x64/lib/ruby/2.1.0/win32/resolv.rb:19:in `get_resolv_info'
      c:/Ruby21-x64/lib/ruby/2.1.0/resolv.rb:969:in `default_config_hash'
      c:/Ruby21-x64/lib/ruby/2.1.0/resolv.rb:986:in `block in lazy_initialize'
      c:/Ruby21-x64/lib/ruby/2.1.0/resolv.rb:979:in `synchronize'
      c:/Ruby21-x64/lib/ruby/2.1.0/resolv.rb:979:in `lazy_initialize'
      c:/Ruby21-x64/lib/ruby/2.1.0/resolv.rb:358:in `block in lazy_initialize'
      c:/Ruby21-x64/lib/ruby/2.1.0/resolv.rb:356:in `synchronize'
      c:/Ruby21-x64/lib/ruby/2.1.0/resolv.rb:356:in `lazy_initialize'
      c:/Ruby21-x64/lib/ruby/2.1.0/resolv.rb:516:in `fetch_resource'
      c:/Ruby21-x64/lib/ruby/2.1.0/resolv.rb:510:in `each_resource'
      c:/Ruby21-x64/lib/ruby/2.1.0/resolv.rb:491:in `getresource'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/remote_fetcher.rb:88:in `api_endpoint'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/source.rb:42:in `api_uri'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/source.rb:170:in `load_specs'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/spec_fetcher.rb:266:in `tuples_for'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/spec_fetcher.rb:226:in `block in available_specs'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/source_list.rb:97:in `each'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/source_list.rb:97:in `each_source'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/spec_fetcher.rb:222:in `available_specs'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/spec_fetcher.rb:102:in `search_for_dependency'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/dependency_installer.rb:216:in `find_gems_with_sourc
es'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/dependency_installer.rb:292:in `find_spec_by_name_and_version'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/dependency_installer.rb:166:in `available_set_for'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/dependency_installer.rb:418:in `resolve_dependencies'
      ,
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/dependency_installer.rb:371:in `install'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/commands/install_command.rb:219:in `install_gem'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/commands/install_command.rb:263:in `block in install_gems'
      _gems'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/commands/install_command.rb:259:in `each'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/commands/install_command.rb:259:in `install_gems'
      c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/commands/install_command.rb:171:in `execute'

```

```
c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/command.rb:305:in `invoke_with_build_args'
c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/command_manager.rb:167:in `process_args'
c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/command_manager.rb:137:in `run'
c:/Ruby21-x64/lib/ruby/2.1.0/rubygems/gem_runner.rb:54:in `run'
c:/Ruby21-x64/bin/gem:21:in `<main>'
```

To resolve this issue I manually modified line 70 of registry.rb:

```
- LOCALE = Encoding.find(Encoding.locale_charmap)
+ LOCALE = Encoding::UTF_8
+ #LOCALE = Encoding.find(Encoding.locale_charmap)
```

Is it possible to change locale_charmap without hacking registry.rb ? UTF-8 is maybe a better default value ?

Thanks,
Guillaume

Associated revisions

Revision ba3da9af7d8db43d421426d506d3f0ab1a0f757e - 12/23/2014 07:03 AM - nobu (Nobuyoshi Nakada)

registry.rb: try en_US message

- ext/win32/lib/win32/registry.rb (Win32::Registry::Error#initialize):
try en_US message if the default message cannot be encoded to
locale. [ruby-core:65295] [Bug #10300]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@48927 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision ba3da9af - 12/23/2014 07:03 AM - nobu (Nobuyoshi Nakada)

registry.rb: try en_US message

- ext/win32/lib/win32/registry.rb (Win32::Registry::Error#initialize):
try en_US message if the default message cannot be encoded to
locale. [ruby-core:65295] [Bug #10300]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@48927 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 802d4f9fb2a8a5f2be1eb6be13dc4f286ad221af - 12/23/2014 08:58 AM - nobu (Nobuyoshi Nakada)

registry.rb: fix buffer overflow

- ext/win32/lib/win32/registry.rb (Win32::Registry::Error#initialize):
should not re-use sliced string as buffer, to get rid of buffer
overflow. [ruby-core:65295] [Bug #10300]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@48928 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 802d4f9f - 12/23/2014 08:58 AM - nobu (Nobuyoshi Nakada)

registry.rb: fix buffer overflow

- ext/win32/lib/win32/registry.rb (Win32::Registry::Error#initialize):
should not re-use sliced string as buffer, to get rid of buffer
overflow. [ruby-core:65295] [Bug #10300]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@48928 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

History

#1 - 09/27/2014 03:07 PM - luislavena (Luis Lavena)

- Subject changed from Troubles installing gems on Windows 7 to Encoding error in conversion from UTF-16LE to UTF-8 to CP850

#2 - 09/27/2014 11:55 PM - duerst (Martin Dürst)

There is no bug in the conversion from (UTF-16LE to) UTF-8 to CP850. CP850 simply doesn't contain U+2019 (RIGHT SINGLE QUOTATION MARK, see <http://www.unicode.org/charts/PDF/U2000.pdf>), see e.g. https://en.wikipedia.org/wiki/Code_page_850. So with the current subject, this bug should actually be rejected.

Then the question is where the U+2019 is coming from. It's rather easy to get one into an otherwise ASCII text file, e.g. with "smart quotes" or some such. The bug is therefore either in the gem (why does a gem called 'asciidoctor' use non-ascii characters :-?), in the RubyGems code, or in the win32/registry code.

#3 - 09/28/2014 02:12 AM - nobu (Nobuyoshi Nakada)

- Status changed from Open to Feedback

Or from FormatMessage?

Can you try with this patch?

```
index 74cc77d..4df59a9 100644
--- a/ext/win32/lib/win32/registry.rb
+++ b/ext/win32/lib/win32/registry.rb
@@ -174,8 +174,15 @@ For detail, see the MSDN[http://msdn.microsoft.com/library/en-us/sysinfo/base/pr
  def initialize(code)
    @code = code
    msg = WCHAR_NUL * 1024
-    len = FormatMessageW.call(0x1200, 0, code, 0, msg, 1024, 0)
-    msg = msg[0, len].encode(LOCALE)
+    lang = 0
+    begin
+      len = FormatMessageW.call(0x1200, 0, code, lang, msg, 1024, 0)
+      msg = msg[0, len].encode(LOCALE)
+    rescue EncodingError
+      raise unless lang == 0
+      lang = 0x0409 # en_US
+      retry
+    end
    super msg.tr("\r".encode(msg.encoding), '').chomp
  end
  attr_reader :code
```

#4 - 09/29/2014 12:06 PM - nanarth (Adrien Bernhardt)

Hello,

I just experienced the same problem than Guillaume on Windows 7 and tried your patch. It solved the problem perfectly.

#5 - 09/29/2014 05:08 PM - naruse (Yui NARUSE)

- Target version set to 2.2.0

#6 - 11/12/2014 09:07 PM - ggrossetie (Guillaume GROSSETIE)

Sorry, I just saw your replies. I will try the patch this week and let you know.

#7 - 11/20/2014 08:19 AM - ggrossetie (Guillaume GROSSETIE)

Nobuyoshi Nakada wrote:

Or from FormatMessage?

Can you try with this patch?

```
index 74cc77d..4df59a9 100644
--- a/ext/win32/lib/win32/registry.rb
+++ b/ext/win32/lib/win32/registry.rb
@@ -174,8 +174,15 @@ For detail, see the MSDN[http://msdn.microsoft.com/library/en-us/sysinfo/base/pr
  def initialize(code)
    @code = code
    msg = WCHAR_NUL * 1024
-    len = FormatMessageW.call(0x1200, 0, code, 0, msg, 1024, 0)
-    msg = msg[0, len].encode(LOCALE)
+    lang = 0
+    begin
+      len = FormatMessageW.call(0x1200, 0, code, lang, msg, 1024, 0)
+      msg = msg[0, len].encode(LOCALE)
+    rescue EncodingError
+      raise unless lang == 0
+      lang = 0x0409 # en_US
+      retry
+    end
    super msg.tr("\r".encode(msg.encoding), '').chomp
  end
  attr_reader :code
```

Thanks that solved the issue!

#8 - 12/23/2014 07:03 AM - nobu (Nobuyoshi Nakada)

- *Status changed from Feedback to Closed*
- *% Done changed from 0 to 100*

Applied in changeset r48927.

registry.rb: try en_US message

- ext/win32/lib/win32/registry.rb (Win32::Registry::Error#initialize):
try en_US message if the default message cannot be encoded to
locale. [\[ruby-core:65295\]](#) [Bug [#10300](#)]

#9 - 12/23/2014 01:34 PM - luislavena (Luis Lavena)

- *Backport changed from 2.0.0: UNKNOWN, 2.1: UNKNOWN to 2.0.0: REQUIRED, 2.1: REQUIRED*

#10 - 01/14/2015 07:22 AM - usa (Usaku NAKAMURA)

- *Backport changed from 2.0.0: REQUIRED, 2.1: REQUIRED to 2.0.0: DONTNEED, 2.1: REQUIRED*