

Ruby trunk - Bug #5570

Encoding of environment variables on Windows

11/04/2011 10:30 PM - now (Nikolai Weibull)

| | | |
|---|--|------------------|
| Status: | Closed | |
| Priority: | Normal | |
| Assignee: | usa (Usaku NAKAMURA) | |
| Target version: | 2.0.0 | |
| ruby -v: | ruby 1.9.3dev (2011-09-13 revision 33263) [i386-mingw32] | Backport: |
| Description | | |
| It seems that the filesystem encoding is forced upon environment variables. There is, however, no guarantee that this encoding is the correct one. Getenv() probably shouldn't be used on Windows at all. Using _wgetenv instead should guarantee that the returned result is encoded properly and it can then be converted to the filesystem encoding. | | |

Associated revisions

Revision fcaee4bd - 03/15/2012 03:57 AM - usa (Usaku NAKAMURA)

- win32/win32.c, include/ruby/win32.h (rb_w32_ugetenv): new API to accept and to return UTF-8 strings.
- win32/win32.c (rb_w32_getenv): follow above change.
- win32/win32.c (rb_w32_get_envIRON): returns UTF-8 environment area.
- hash.c (env_str_new, rb_f_getenv, env_fetch): follow above changes. [Bug #5570] [ruby-core:40737]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@35030 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 35030 - 03/15/2012 03:57 AM - usa (Usaku NAKAMURA)

- win32/win32.c, include/ruby/win32.h (rb_w32_ugetenv): new API to accept and to return UTF-8 strings.
- win32/win32.c (rb_w32_getenv): follow above change.
- win32/win32.c (rb_w32_get_envIRON): returns UTF-8 environment area.
- hash.c (env_str_new, rb_f_getenv, env_fetch): follow above changes. [Bug #5570] [ruby-core:40737]

Revision 35030 - 03/15/2012 03:57 AM - usa (Usaku NAKAMURA)

- win32/win32.c, include/ruby/win32.h (rb_w32_ugetenv): new API to accept and to return UTF-8 strings.
- win32/win32.c (rb_w32_getenv): follow above change.
- win32/win32.c (rb_w32_get_envIRON): returns UTF-8 environment area.
- hash.c (env_str_new, rb_f_getenv, env_fetch): follow above changes. [Bug #5570] [ruby-core:40737]

Revision 35030 - 03/15/2012 03:57 AM - usa (Usaku NAKAMURA)

- win32/win32.c, include/ruby/win32.h (rb_w32_ugetenv): new API to accept and to return UTF-8 strings.
- win32/win32.c (rb_w32_getenv): follow above change.
- win32/win32.c (rb_w32_get_environ): returns UTF-8 environment area.
- hash.c (env_str_new, rb_f_getenv, env_fetch): follow above changes. [Bug #5570] [ruby-core:40737]

Revision 35030 - 03/15/2012 03:57 AM - usa (Usaku NAKAMURA)

- win32/win32.c, include/ruby/win32.h (rb_w32_ugetenv): new API to accept and to return UTF-8 strings.
- win32/win32.c (rb_w32_getenv): follow above change.
- win32/win32.c (rb_w32_get_environ): returns UTF-8 environment area.
- hash.c (env_str_new, rb_f_getenv, env_fetch): follow above changes. [Bug #5570] [ruby-core:40737]

Revision 35030 - 03/15/2012 03:57 AM - usa (Usaku NAKAMURA)

- win32/win32.c, include/ruby/win32.h (rb_w32_ugetenv): new API to accept and to return UTF-8 strings.
- win32/win32.c (rb_w32_getenv): follow above change.
- win32/win32.c (rb_w32_get_environ): returns UTF-8 environment area.
- hash.c (env_str_new, rb_f_getenv, env_fetch): follow above changes. [Bug #5570] [ruby-core:40737]

Revision 35030 - 03/15/2012 03:57 AM - usa (Usaku NAKAMURA)

- win32/win32.c, include/ruby/win32.h (rb_w32_ugetenv): new API to accept and to return UTF-8 strings.
- win32/win32.c (rb_w32_getenv): follow above change.
- win32/win32.c (rb_w32_get_environ): returns UTF-8 environment area.
- hash.c (env_str_new, rb_f_getenv, env_fetch): follow above changes. [Bug #5570] [ruby-core:40737]

History

#1 - 03/11/2012 04:43 PM - ko1 (Koichi Sasada)

- Assignee set to usa (Usaku NAKAMURA)

#2 - 03/11/2012 11:36 PM - naruse (Yui NARUSE)

- Status changed from Open to Rejected

Ruby uses AreFileApisANSI() ? GetACP() : GetOEMCP() on Windows, not environment variables.

#3 - 03/11/2012 11:59 PM - now (Nikolai Weibull)

On Sun, Mar 11, 2012 at 15:36, Yui NARUSE naruse@airemix.jp wrote:

Issue [#5570](#) has been updated by Yui NARUSE.

Status changed from Open to Rejected

Ruby uses AreFileApisANSI() ? GetACP() : GetOEMCP() on Windows, not environment variables.

The problem is with the encoding of the values of the environment variables, not whether environment variables are used to determine the encoding of file names or not.

Please re-open.

#4 - 03/12/2012 12:05 AM - naruse (Yui NARUSE)

- Status changed from Rejected to Assigned

Ah, you are saying "use Wide API".

#5 - 03/12/2012 12:10 AM - luislavena (Luis Lavena)

Correct,

A example that reproduces the issue:

```
C:\Users\Luis>ruby -v
ruby 1.9.3p155 (2012-03-03 revision 34878) [i386-mingw32]
```

```
C:\Users\Luis>chcp
Active code page: 1252
```

```
C:\Users\Luis>SET FOO=fóñè
```

```
C:\Users\Luis>ruby -e "puts ENV['FOO']"
fóñè
```

```
C:\Users\Luis>chcp 65001
Active code page: 65001
```

```
C:\Users\Luis>ruby -e "puts ENV['FOO']"
f◆◆◆
```

This also happens with environment variables used to construct users home directory "~" if they contain accented characters.

#6 - 03/12/2012 06:44 AM - nobu (Nobuyoshi Nakada)

- Status changed from Assigned to Feedback

- Target version set to 2.0.0

Which codepage should be used there, ACP or OEMCP?

#7 - 03/12/2012 04:53 PM - now (Nikolai Weibull)

On Sun, Mar 11, 2012 at 22:44, Nobuyoshi Nakada nobu@ruby-lang.org wrote:

Issue [#5570](#) has been updated by Nobuyoshi Nakada.

Status changed from Assigned to Feedback

Target version set to 2.0.0

Which codepage should be used there, ACP or OEMCP?

I'm not sure that I understand what the question is referring to, but the encoding of the value of environment variables should be the

filesystem encoding on Windows, just as it is on Unix. The only difference is that `_wgetenv` should be used instead of `getenv` so that the transcoding can be done reliably.

#8 - 03/15/2012 01:23 PM - now (Nikolai Weibull)

On Thu, Mar 15, 2012 at 02:51, U.Nakamura usa@garbagecollect.jp wrote:

Hello,

In message "[ruby-core:43237] [ruby-trunk - Bug #5570][Feedback] Encoding of environment variables on Windows" on Mar.12,2012 06:44:04, nobu@ruby-lang.org wrote:

Which codepage should be used there, ACP or OEMCP?

maybe locale?

Huh? Again, this isn't the issue.

#9 - 03/16/2012 12:46 PM - usa (Usaku NAKAMURA)

- Status changed from Feedback to Closed

- % Done changed from 0 to 100

This issue was solved with changeset r35030.

Nikolai, thank you for reporting this issue.

Your contribution to Ruby is greatly appreciated.

May Ruby be with you.

-
- win32/win32.c, include/ruby/win32.h (rb_w32_ugetenv): new API to accept and to return UTF-8 strings.
 - win32/win32.c (rb_w32_getenv): follow above change.
 - win32/win32.c (rb_w32_get_envIRON): returns UTF-8 environment area.
 - hash.c (env_str_new, rb_f_getenv, env_fetch): follow above changes. [Bug #5570] [ruby-core:40737]

#10 - 03/16/2012 02:23 PM - now (Nikolai Weibull)

On Thu, Mar 15, 2012 at 05:16, Nikolai Weibull now@bitwi.se wrote:

On Thu, Mar 15, 2012 at 02:51, U.Nakamura usa@garbagecollect.jp wrote:

Hello,

In message "[ruby-core:43237] [ruby-trunk - Bug #5570][Feedback] Encoding of environment variables on Windows" on Mar.12,2012 06:44:04, nobu@ruby-lang.org wrote:

Which codepage should be used there, ACP or OEMCP?

maybe locale?

Huh? Again, this isn't the issue.

This seems to have been resolved in 35030.

Thanks!