

Ruby master - Bug #4097

Unexpected result of STDIN.read on Windows

11/29/2010 06:26 PM - phasis68 (Heesob Park)

Status: Third Party's Issue	
Priority: Normal	
Assignee: tarui (Masaya Tarui)	
Target version: 1.9.3	
ruby -v: ruby 1.9.3dev (2010-11-28 trunk 29965) [i386-mswin32_90]	Backport:
Description =begin On Ruby 1.9.x, in case of non-ASCII input, STDIN.read(n) returns some garbage attached string. C:\work>ruby -ve 'a=STDIN.read(10);p a;p a.length' ruby 1.9.3dev (2010-11-28 trunk 29965) [i386-mswin32_90] [] [] [] abcd "\xB0\xA1\xB3\xAA\xB4\xD9\xB6\xF3ab\x00\x00\xB8t" 14 On the other hand, Ruby 1.8.6 works fine. C:\work>ruby -ve 'a=STDIN.read(10);p a;p a.length' ruby 1.8.6 (2010-02-04 patchlevel 398) [i386-mingw32] [] [] [] abcd "\260\241\263\252\264\331\266\363ab" 10 =end	

History

#1 - 11/29/2010 08:44 PM - usa (Usaku NAKAMURA)

=begin
Hello,

In message "[ruby-core:33460] [Ruby 1.9-Bug#4097][Open] Unexpected result of STDIN.read on Windows"
on Nov.29,2010 18:26:13, redmine@ruby-lang.org wrote:

On Ruby 1.9.x, in case of non-ASCII input, STDIN.read(n) returns some garbage attached string.

What version of Windows do you use?
I guess you use Korean version of 32bit XP, don't you?

Tarui-san tested many cases on Japanese version of 32bit XP,
and has found that this seems to be a bug of Windows itself...

Regards,
--
U.Nakamura usa@garbagecollect.jp

=end

#2 - 11/29/2010 09:45 PM - luislavena (Luis Lavena)

=begin
On Mon, Nov 29, 2010 at 8:44 AM, U.Nakamura usa@garbagecollect.jp wrote:

Hello,

In message "[ruby-core:33460] [Ruby 1.9-Bug#4097][Open] Unexpected result of STDIN.read on Windows"
on Nov.29,2010 18:26:13, redmine@ruby-lang.org wrote:

On Ruby 1.9.x, in case of non-ASCII input, STDIN.read(n) returns some garbage attached string.


```
C:\work>ruby -ve 'a=STDIN.read(10);p a;p a.length'
ruby 1.9.3dev (2010-11-28 trunk 29965) [i386-mswin32_90]
[] [] []abcd
"\xB0\xA1\xB3\xAA\xB4\xD9\xB6\xF3ab\x00\x00\xB8t"
14
```

Regards,
Masaya TARUI

=end

#5 - 11/30/2010 12:07 PM - phasis68 (Heesob Park)

=begin
Hi,

2010/11/30 Masaya TARUI tarui@prx.jp:

Hello,

WindowsXP seems have a bug at read functions under multibyte console inputs.
I found a issue of coming from same bug of Windows. :-(

does anybody have a good workaround idea ?

```
ruby -ve 'a=STDIN.read(6);p [a,a.length];a=STDIN.read(2);p [a,a.length];'
ruby 1.9.3dev (2010-11-30 trunk 29978) [i386-mswin32_100]
[] [] [] []aiueo
["\x82\xA0\x82\xA2\x82\xA4", 6]
["iu", 2]
```

On Ruby 1.9.x, in case of non-ASCII input, STDIN.read(n) returns some garbage attached string.

```
C:\work>ruby -ve 'a=STDIN.read(10);p a;p a.length'
ruby 1.9.3dev (2010-11-28 trunk 29965) [i386-mswin32_90]
[] [] []abcd
"\xB0\xA1\xB3\xAA\xB4\xD9\xB6\xF3ab\x00\x00\xB8t"
14
```

I found ReadFile on console reads data per charachacter not byte.

Here is a workaround patch.

```
--- win32.c 2010-11-30 12:02:33.000000000 +0900
+++ win32.c.new 2010-11-30 12:01:46.000000000 +0900
@@ -5091,6 +5091,34 @@
pol = &ol;
}
```

- if (is_console(_osfhnd(fd)) && len!=16384) {
- int len2=0;
- while(len2<len) {
- if (!ReadFile((HANDLE)_osfhnd(fd), buf, 1, &read, pol)) {
- err = GetLastError();
- if (err != ERROR_IO_PENDING) {
- if (pol) CloseHandle(ol.hEvent);
- if (err == ERROR_ACCESS_DENIED)
- errno = EBADF;
- else if (err == ERROR_BROKEN_PIPE || err == ERROR_HANDLE_EOF) { + MTHREAD_ONLY(LeaveCriticalSection(&_pioinfo(fd)->lock));
- return 0;
- }
- else
- errno = map_errno(err); + MTHREAD_ONLY(LeaveCriticalSection(&_pioinfo(fd)->lock));
- return -1;
- }
- }
- len2 += read;
- buf = (char *)buf + read;
- }
- ret += len;
- if (size > 0)
- goto retry; +

```
• } else { if (!ReadFile((HANDLE)_osfhd(fd), buf, len, &read, pol)) { err = GetLastError(); if (err != ERROR_IO_PENDING) { @@ -5154,6
+5182,7 @@ if (size > 0) goto retry; }
```

```
• }
```

```
MTHREAD_ONLY(LeaveCriticalSection(&_pioinfo(fd)->lock));
```

Regards,
Park Heesob

=end

#6 - 12/01/2010 11:08 AM - usa (Usaku NAKAMURA)

- Status changed from Open to Assigned

- Assignee set to tarui (Masaya Tarui)

=begin

=end

#7 - 06/26/2011 11:21 PM - nahi (Hiroshi Nakamura)

- Target version changed from 2.0.0 to 1.9.3

#8 - 07/21/2011 02:03 PM - kosaki (Motohiro KOSAKI)

Tarui-san, ping?

#9 - 07/27/2011 04:34 AM - tarui (Masaya Tarui)

- Status changed from Assigned to Third Party's Issue

Sorry for a delayed response.

Now, STDIN.read(n) under multibyte console inputs might return n+1 bytes String.(by r29980 and r30280)
Multibyte character is never split in read of MS runtime.

And, it is difficult to do STDIN.ungetc last byte because of lapping C-level read function.

I think that

- 1) it's windows bug,
- 2) we don't have an api base workaround ,
- and
- 3) we can apply a workaround to application.

So, I change status to 3rd party's issue.

However, the patch is always a welcome.

Thanks,
Masaya TARUI