

Ruby - Bug #4491

Some methods of Dir/File do not encode dirname/filename transparently

03/10/2011 08:22 PM - yimutang (Joey Zhou)

Status:	Closed	Backport:
Priority:	Normal	
Assignee:	usa (Usaku NAKAMURA)	
Target version:		
ruby -v:	ruby 1.9.2p180 (2011-02-18) [i386-mingw32]	
Description		
<pre>=begin My Ruby version is: ruby 1.9.2p180 (2011-02-18) [i386-mingw32]</pre>		
<p>My OS is Windows 7 (Simplified Chinese), its cmd.exe has a default codepage of CP936 (aka GBK), but I want my scripts to be written in UTF-8, so I've set the scripts' to UTF-8 and add "# encoding: utf-8" on the first line.</p>		
<p>I find when I open a file, the filename will be encoded to the system's locale charmap automatically:</p>		
<pre>puts File.open("UTF8file").readline #this will be fine, although the "UTF8file" literal is UTF-8 puts File.open("UTF8file".encode('utf-8')).readline #explicitly encoded puts File.open("UTF8file".encode('gbk')).readline puts File.open("UTF8file".encode('big5')).readline puts File.open("UTF8file".encode('shift_jis')).readline #all fine, all encoded to Encoding.locale_charmap before asking the OS for the file, no matter to what it's encoded here</pre>		
<p>I have tested many methods of File & Dir class, most of them have this intimate and convenient feature, but some are not:</p>		
<pre>p Dir.entries("UTF8dir".encode('gbk')) # explicit encode is required, or "No such file or directory" error. Since my local charmap is gbk, it can not be encoded to utf-8, big5, shift_jis p Dir.glob("UTF8dir/*.encode('gbk')) Dir.foreach("UTF8dir".encode('gbk')) { f puts f} Dir.open("UTF8dir".encode('gbk')).each { f puts f} Dir.new("UTF8dir".encode('gbk')).each { f puts f} # all must explicitly encoded to locale_charmap (my situation is GBK)</pre>		
<p>and a File method require encoding too:</p>		
<pre>puts File.absolute_path("UTF8file".encode('gbk'))</pre>		
<p>These methods may be missed, they may should act the same way as Dir.chdir("UTF8dir") or File.open("UTF8dir"), with a feature of transparently encoding.</p>		
<p>Best regards.</p>		
<p>Joey =end</p>		

Associated revisions

Revision 22d5cf92 - 04/28/2011 07:22 AM - U.Nakamura

- win32/{win32.c,dir.h} (rb_w32_uopendir): new API to pass UTF-8 path.
- win32/win32.c (opendir_internal, rb_w32_opendir): extract and merge common part of rb_w32_opendir() and rb_w32_uopendir().
- dir.c (do_opendir, glob_helper): encoding.
- dir.c (dir_initialize, do_opendir): convert path to UTF-8 and call rb_w32_uopendir() instead of rb_w32_opendir() on Windows. fixes #4491, reported by Joey Zhou.

History

#1 - 04/26/2011 02:51 PM - yimutang (Joey Zhou)

```
=begin
Is this issue forgotten? :)
=end
```

#2 - 04/26/2011 06:13 PM - naruse (Yui NARUSE)

- *Category set to M17N*
- *Status changed from Open to Assigned*
- *Assignee set to naruse (Yui NARUSE)*

```
=begin
=end
```

#3 - 04/28/2011 02:15 PM - naruse (Yui NARUSE)

- *Assignee changed from naruse (Yui NARUSE) to usa (Usaku NAKAMURA)*

```
=begin
=end
```

#4 - 04/28/2011 02:53 PM - usa (Usaku NAKAMURA)

```
=begin
We now consider that this is a bug.

Current code is too complex, so to judge the reason why somebody wrote such code is difficult.
We want to get time a little more, sorry.

=end
```

#5 - 04/28/2011 04:22 PM - usa (Usaku NAKAMURA)

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

```
=begin
This issue was solved with changeset r31372.
Joey, thank you for reporting this issue.
Your contribution to Ruby is greatly appreciated.
May Ruby be with you.
```

-
- win32/{win32.c,dir.h} (rb_w32_uopendir): new API to pass UTF-8 path.
 - win32/win32.c (opendir_internal, rb_w32_opendir): extract and merge common part of rb_w32_opendir() and rb_w32_uopendir().
 - dir.c (do_opendir, glob_helper): encoding.
 - dir.c (dir_initialize, do_opendir): convert path to UTF-8 and call rb_w32_uopendir() instead of rb_w32_opendir() on Windows. fixes [#4491](#), reported by Joey Zhou.
- ```
=end
```