

Ruby master - Bug #14456

Dir.glob with FNM_CASEFOLD gives ArgumentError: invalid byte sequence in UTF-8

02/06/2018 01:43 PM - Gondolin (Damien Robert)

Status: Open	
Priority: Normal	
Assignee:	
Target version:	
ruby -v: 2.5.0p0	Backport: 2.3: UNKNOWN, 2.4: UNKNOWN, 2.5: UNKNOWN
Description With ruby 2.5.0p0, in a folder that contains a file encoded in latin-1, I get the following error: <pre>Dir.glob('*a', File::FNM_CASEFOLD) ArgumentError: invalid byte sequence in UTF-8</pre> Note that Dir.glob(*, File::FNM_CASEFOLD), Dir.glob('a*', File::FNM_CASEFOLD) and Dir.glob('*a') work, so it is a bit strange that Dir.glob('*a', File::FNM_CASEFOLD) does not.	
Related issues: Has duplicate Ruby master - Bug #14455: Dir.glob with FNM_CASEFOLD gives Argu... Closed	

History

#1 - 02/06/2018 01:48 PM - Gondolin (Damien Robert)

Apologies for the spam, my browser submitted this bug several times and I don't know how to erase the other bug reports.

#2 - 02/06/2018 02:05 PM - shevegen (Robert A. Heiler)

I don't know how to erase the other bug reports.

Not sure if it can be removed, but I think you can change the status to "closed" (or someone from the ruby core team probably could).

To the bug report, just out of curiosity, can you avoid the UTF problem if you change to use another encoding before calling Dir.glob? For example, I usually use 'ISO-8859-1' mostly due to german umlauts but the terminal/shell used, such as my case usually mate-terminal (based on vte) these days, there is also an option where I can decide to use UTF-8 as locale, or the current active locale (ISO-8859-1). When I use both ISO and the locale setting, I almost never get invalid byte sequence errors (and of course if I sync any external input read... File.read() has an option for specifying the :encoding, perhaps Dir.glob() could also benefit from a hash with a key called :encoding but I am perhaps digressing...)

#3 - 02/07/2018 12:55 AM - duerst (Martin Dürst)

- Has duplicate Bug #14455: Dir.glob with FNM_CASEFOLD gives ArgumentError: invalid byte sequence in UTF-8 added

#4 - 02/07/2018 08:33 PM - Gondolin (Damien Robert)

shevegen (Robert A. Heiler) wrote:

Not sure if it can be removed, but I think you can change the status to "closed" (or someone from the ruby core team probably could).

Someone did, thanks!

To the bug report, just out of curiosity, can you avoid the UTF problem if you change to use another encoding before calling Dir.glob?

Yes: `Dir.glob("**a".encode!("ISO-8859-1"), File::FNM_CASEFOLD)` works.

#5 - 02/07/2018 08:36 PM - Gondolin (Damien Robert)

- *Description updated*