

Ruby - Bug #17280

Dir.glob with FNM_DOTMATCH matches "." and ".." and results in duplicated entries

10/22/2020 02:20 PM - Eregon (Benoit Daloze)

Status:	Closed	
Priority:	Normal	
Assignee:		
Target version:		
ruby -v:	ruby 2.6.6p146 (2020-03-31 revision 67876) [x86_64-linux]	Backport: 2.5: UNKNOWN, 2.6: UNKNOWN, 2.7: UNKNOWN
Description <pre>% ruby -e 'p Dir.glob("**/*", File::FNM_DOTMATCH) ' [".", "bar", "bar/.", "bar/.baz", "bar/.baz/.", "bar/.baz/qux"] % ruby -e 'p Dir.glob("**", File::FNM_DOTMATCH) ' [".", "..", "bar"] % ruby -e 'p Dir.glob("*", File::FNM_DOTMATCH) ' [".", "..", "bar"]</pre> I think "." was never intended by the user here, is it a bug? Not sure about "..". Note it also causes duplicated entries: bar and bar/.baz are twice in the Array! I think .. should always be ignored for glob purposes, it escapes the current directory. And . seems useless and causing duplicates. I think the intention of users of File::FNM_DOTMATCH is to match file/directories starting with a . like .baz. Probably Dir.glob("**/{*,.}") is a safer way to achieve that, but still I think FNM_DOTMATCH should not produce such weird results. From https://github.com/oracle/truffleruby/issues/2116 I could not figure out what was the intended semantics for FNM_DOTMATCH with regards to . and ...		
Related issues:		
Related to Ruby - Bug #16831: Running `Pathname#glob` with `File::FNM_DOTMATCH`...		Closed
Related to Ruby - Bug #17283: Why does Dir.glob's ** match files in current d...		Closed
Related to Ruby - Bug #18436: Fix Pathname dot directory globbing		Closed

Associated revisions

Revision 7dc0511ea4be210f82abb1c82a31aec3a4fe5736 - 01/12/2021 11:02 AM - nobu (Nobuyoshi Nakada)

Remove "." and ".." from Dir.glob with FNM_DOTMATCH [Bug #17280]

Co-authored-by: Jeremy Evans code@jeremyevans.net

Revision 7dc0511ea4be210f82abb1c82a31aec3a4fe5736 - 01/12/2021 11:02 AM - nobu (Nobuyoshi Nakada)

Remove "." and ".." from Dir.glob with FNM_DOTMATCH [Bug #17280]

Co-authored-by: Jeremy Evans code@jeremyevans.net

Revision 7dc0511e - 01/12/2021 11:02 AM - nobu (Nobuyoshi Nakada)

Remove "." and ".." from Dir.glob with FNM_DOTMATCH [Bug #17280]

Co-authored-by: Jeremy Evans code@jeremyevans.net

History

#1 - 10/22/2020 02:21 PM - Eregon (Benoit Daloze)

- Related to Bug #16831: Running `Pathname#glob` with `File::FNM\_DOTMATCH` option loses `.` and `..` added

#2 - 10/25/2020 12:16 PM - Eregon (Benoit Daloze)

- Related to Bug #17283: Why does Dir.glob's ** match files in current directory? added

#3 - 11/19/2020 12:26 AM - jeremyevans0 (Jeremy Evans)

So there is two separate issues you are discussing. One issue is when the recursive glob is used ("**/*") where the same folder shows up under two paths, once in the parent directory and once in its own directory. The second issue is when a non-recursive glob is used that the reference to the parent folder (..) shows up (this is always excluded in recursive mode, even without FNM_DOTMATCH).

The first issue we can solve by skipping the current directory entry if it is . and we are in recursive mode and the current path (parent directory) matches the previous entry in the resulting array. This approach works as glob uses a depth-first search and not a breadth-first search, as long as . is the first entry in the directory.

The second issue we can solve by not matching .. entry if the glob is magical == 2 (* sets that).

I added a pull request with possible fixes: <https://github.com/ruby/ruby/pull/3789>. We'll see if it passes CI. Even if so, this changes behavior and I'm not sure I consider the current behavior in either case a bug, so we should get feedback from more committers as to what the expected behavior is.

#4 - 11/19/2020 12:53 PM - Eregon (Benoit Daloze)

Thanks, that looks good to me.

I think everyone expects FNM_DOTMATCH to match dotfiles, and . and .. are not dotfiles.

Or is there is some other purpose for FNM_DOTMATCH?

#5 - 11/19/2020 12:58 PM - Eregon (Benoit Daloze)

I wonder if we could simplify the logic to never include .. for Dir.glob, and only include . for the initial directory and nowhere else. Are there cases where .. and . would ever be wanted, besides Dir.glob(".") => ["."]?

Not sure it's even a good idea to include . for Dir.glob("", File::FNM_DOTMATCH), the intention of * seems clearly to list all files/subdirs under this directory, not this directory itself, and certainly not the parent directory.

#6 - 11/23/2020 02:36 PM - Eregon (Benoit Daloze)

Currently, even Dir[".*"] and Dir["{*,.*}"] match .., which seems surprising. The PR fixes those, and I think the resulting behavior is much more useful.

#7 - 01/12/2021 07:38 AM - mame (Yusuke Endoh)

https://github.com/ruby/dev-meeting-log/blob/master/DevelopersMeeting20201210Japan.md#bug-17280-dirglob-with-fnm_dotmatch-matches--and--and-results-in-duplicated-entries-jeremyevans0

matz: try to remove "." and ".." on 3.1 (early timing to check compatibility issues)

I think we can progress this change.

#8 - 01/12/2021 11:03 AM - nobu (Nobuyoshi Nakada)

- Status changed from Open to Closed

Applied in changeset [git|7dc0511ea4be210f82abb1c82a31aec3a4fe5736](https://github.com/ruby/ruby/commit/7dc0511ea4be210f82abb1c82a31aec3a4fe5736).

Remove "." and ".." from Dir.glob with FNM_DOTMATCH [Bug #17280]

Co-authored-by: Jeremy Evans code@jeremyevans.net

#9 - 12/27/2021 12:09 AM - znz (Kazuhiro NISHIYAMA)

- Related to Bug #18436: Fix Pathname dot directory globbing added