

Ruby - Bug #19485

Unexpected behavior in squiggly heredocs

03/08/2023 03:24 PM - jemmai (Jemma Issroff)

Status:	Closed	Backport: 2.7: UNKNOWN, 3.0: REQUIRED, 3.1: DONE, 3.2: DONE
Priority:	Normal	
Assignee:	core	
Target version:		
ruby -v:	3.2.1	
Description		
<p>Based on the squiggly heredoc documentation, I found the following to be unexpected behavior. Explicitly, the documentation specifies, "The indentation of the least-indented line will be removed from each line of the content."</p> <p>After running:</p> <pre>File.write("test.rb", "p <<~EOF\n\tA\n b\nEOF\n")</pre> <p>and then ruby test.rb, I get the following output:</p> <pre>"\tA\nb\n"</pre> <p>The least-indented line above is b, however, no leading whitespace is removed from the line containing \tA.</p> <p>For another example:</p> <pre>File.write("test.rb", "p <<~EOF\n\tA\n \tB\nEOF\n")</pre> <p>ruby test.rb gives:</p> <pre>"A\nB\n"</pre> <p>In this case, the \t was removed from the line containing A, but more whitespace than that (\t) was removed from the line containing B.</p> <p>After seeing the first example, I assumed that the documentation was out of date, and that I should fix it to read that \t would never be converted into space characters in order to remove leading whitespace. But after the second example, it seems like this is a bug in removing leading whitespace.</p> <p>Can someone please explain what the rules should be on squiggly heredocs? I can implement a fix to adhere to the rules, or can update the documentation, I am just unsure of what the rules should be because the above two examples reflect unexpected behavior in two distinct ways.</p>		

Associated revisions

Revision e7342e76dfd26237c604e42f9a59a1eaa578c94e - 03/19/2023 09:10 AM - nobu (Nobuyoshi Nakada)

[Bug #19485] [DOC] Mention tabs in indentation of heredoc identifier

Co-Authored-By: sawa (Tsuyoshi Sawada) sawadatsuyoshi@gmail.com

Revision e7342e76dfd26237c604e42f9a59a1eaa578c94e - 03/19/2023 09:10 AM - nobu (Nobuyoshi Nakada)

[Bug #19485] [DOC] Mention tabs in indentation of heredoc identifier

Co-Authored-By: sawa (Tsuyoshi Sawada) sawadatsuyoshi@gmail.com

Revision e7342e76 - 03/19/2023 09:10 AM - nobu (Nobuyoshi Nakada)

[Bug #19485] [DOC] Mention tabs in indentation of heredoc identifier

Co-Authored-By: sawa (Tsuyoshi Sawada) sawadatsuyoshi@gmail.com

Revision b93e2223300bc54dfa387ffb9fa3d48ecbe670f0 - 03/23/2023 03:04 AM - naruse (Yui NARUSE)

```
merge revision(s) e7342e76dfd26237c604e42f9a59a1eaa578c94e: [Backport #19485]

[Bug #19485] [DOC] Mention tabs in indentation of heredoc identifier

Co-Authored-By: sawa (Tsuyoshi Sawada) <sawadatsuyoshi@gmail.com>
---
doc/syntax/literals.rdoc | 6 ++++++
1 file changed, 6 insertions(+)
```

Revision b93e2223300bc54dfa387ffb9fa3d48ecbe670f0 - 03/23/2023 03:04 AM - naruse (Yui NARUSE)

```
merge revision(s) e7342e76dfd26237c604e42f9a59a1eaa578c94e: [Backport #19485]

[Bug #19485] [DOC] Mention tabs in indentation of heredoc identifier

Co-Authored-By: sawa (Tsuyoshi Sawada) <sawadatsuyoshi@gmail.com>
---
doc/syntax/literals.rdoc | 6 ++++++
1 file changed, 6 insertions(+)
```

Revision b93e2223 - 03/23/2023 03:04 AM - naruse (Yui NARUSE)

```
merge revision(s) e7342e76dfd26237c604e42f9a59a1eaa578c94e: [Backport #19485]

[Bug #19485] [DOC] Mention tabs in indentation of heredoc identifier

Co-Authored-By: sawa (Tsuyoshi Sawada) <sawadatsuyoshi@gmail.com>
---
doc/syntax/literals.rdoc | 6 ++++++
1 file changed, 6 insertions(+)
```

Revision 19af12ff195aba64bdca7a83f564f2c0e46061c0 - 03/26/2023 04:12 AM - nagachika (Tomoyuki Chikanaga)

```
merge revision(s) e7342e76dfd26237c604e42f9a59a1eaa578c94e: [Backport #19485]

[Bug #19485] [DOC] Mention tabs in indentation of heredoc identifier

Co-Authored-By: sawa (Tsuyoshi Sawada) <sawadatsuyoshi@gmail.com>
---
doc/syntax/literals.rdoc | 6 ++++++
1 file changed, 6 insertions(+)
```

Revision 19af12ff195aba64bdca7a83f564f2c0e46061c0 - 03/26/2023 04:12 AM - nagachika (Tomoyuki Chikanaga)

```
merge revision(s) e7342e76dfd26237c604e42f9a59a1eaa578c94e: [Backport #19485]

[Bug #19485] [DOC] Mention tabs in indentation of heredoc identifier

Co-Authored-By: sawa (Tsuyoshi Sawada) <sawadatsuyoshi@gmail.com>
---
doc/syntax/literals.rdoc | 6 ++++++
1 file changed, 6 insertions(+)
```

Revision 19af12ff - 03/26/2023 04:12 AM - nagachika (Tomoyuki Chikanaga)

```
merge revision(s) e7342e76dfd26237c604e42f9a59a1eaa578c94e: [Backport #19485]

[Bug #19485] [DOC] Mention tabs in indentation of heredoc identifier

Co-Authored-By: sawa (Tsuyoshi Sawada) <sawadatsuyoshi@gmail.com>
---
doc/syntax/literals.rdoc | 6 ++++++
1 file changed, 6 insertions(+)
```

History

#1 - 03/13/2023 12:52 PM - Dan0042 (Daniel DeLorme)

I think what's happening here is that tabs are not converted directly to 8 spaces, but to "move ahead to next multiple of 8 chars". So in that sense "\t" and " \t" are equivalent. It's the same behavior as 10.times{ |i| print " *i","\t",i,"\n" }

#2 - 03/15/2023 05:16 AM - nobu (Nobuyoshi Nakada)

- Status changed from Open to Assigned
- Assignee set to core

- Backport changed from 2.7: UNKNOWN, 3.0: UNKNOWN, 3.1: UNKNOWN, 3.2: UNKNOWN to 2.7: UNKNOWN, 3.0: REQUIRED, 3.1: REQUIRED, 3.2: REQUIRED

My [draft](#) is:

Note that the "indentation" is counted like as each horizontal tabs are expanded to spaces up to the next tab stop column (per 8 columns), and each indentation to be removed is the longest tabs and spaces sequence where the next column does not exceed the least-indentation.

Does this make sense?

#3 - 03/15/2023 07:57 AM - sawa (Tsuyoshi Sawada)

nobu (Nobuyoshi Nakada) wrote in [#note-2](#):

My [draft] is:

Note that the "indentation" is counted like as each horizontal tabs are expanded to spaces up to the next tab stop column (per 8 columns), and each indentation to be removed is the longest tabs and spaces sequence where the next column does not exceed the least-indentation.

I find the sentence too long and a little too difficult to parse/understand. What about something like this:

For the purpose of measuring an indentation, a horizontal tab is regarded as a sequence of one to eight spaces such that the column position corresponding to its end is a multiple of eight. The amount to be removed is counted in terms of the number of spaces. If the boundary appears in the middle of a tab, that tab is not removed.

#4 - 03/15/2023 08:40 AM - ioquatix (Samuel Williams)

I don't think it's a good idea to assume a tab is 8 spaces.

Regarding indentation, it might be a nice simplification to only consider the first line in the squiggly heredoc. That's what I've done in the past - it's predictable and easy to explain.

i.e.

```
x = <<~FOO
  1
  2
  3
  FOO
```

At most 4 spaces is removed from each line. The first line determines this. Anyway, maybe it's irrelevant to this discussion. But that's how I've implemented it in my own language/interpreter in the past.

Python also has the idea of consistent indentation.

That means mixed spaces/tabs are not considered the same. If someone indents with "SSSSTT" and "TT" on two lines, it's considered invalid and/or not removed. Since you can't determine the equivalence of "S" (space) and "T" (tab) characters. Assuming there is a mapping from tabs to spaces is incorrect IMHO.

#5 - 03/15/2023 11:28 AM - Eregon (Benoit Daloze)

Another condition could be only accept tabs in squiggly heredoc if they prefix all lines of the squiggly heredoc? (otherwise SyntaxError, including for the 2 cases in the description)

(I wish tabs would just not be accepted as indentation for Ruby, but well that's probably a pointless discussion, even though it seems 99% of the community agrees there)

#6 - 03/15/2023 03:07 PM - jemmai (Jemma Issroff)

sawa (Tsuyoshi Sawada) wrote in [#note-3](#):

For the purpose of measuring an indentation, a horizontal tab is regarded as a sequence of one to eight spaces such that the column position corresponding to its end is a multiple of eight. The amount to be removed is counted in terms of the number of spaces. If the boundary appears in the middle of a tab, that tab is not removed.

This documentation is very clear to me, and explains both cases I've mentioned in a way that is easy to understand.

#7 - 03/19/2023 09:51 AM - nobu (Nobuyoshi Nakada)

- Status changed from Assigned to Closed

Applied in changeset [git|e7342e76dfd26237c604e42f9a59a1eaa578c94e](https://github.com/ruby/ruby/commit/e7342e76dfd26237c604e42f9a59a1eaa578c94e).

[Bug [#19485](#)] [DOC] Mention tabs in indentation of heredoc identifier

Co-Authored-By: sawa (Tsuyoshi Sawada) sawadatsuyoshi@gmail.com

#8 - 03/23/2023 03:04 AM - naruse (Yui NARUSE)

- Backport changed from 2.7: UNKNOWN, 3.0: REQUIRED, 3.1: REQUIRED, 3.2: REQUIRED to 2.7: UNKNOWN, 3.0: REQUIRED, 3.1: REQUIRED, 3.2: DONE

ruby_3_2 b93e2223300bc54dfa387ffb9fa3d48ecbe670f0 merged revision(s) e7342e76dfd26237c604e42f9a59a1eaa578c94e.

#9 - 03/26/2023 04:21 AM - nagachika (Tomoyuki Chikanaga)

- Backport changed from 2.7: UNKNOWN, 3.0: REQUIRED, 3.1: REQUIRED, 3.2: DONE to 2.7: UNKNOWN, 3.0: REQUIRED, 3.1: DONE, 3.2: DONE

ruby_3_1 19af12ff195aba64bdca7a83f564f2c0e46061c0 merged revision(s) e7342e76dfd26237c604e42f9a59a1eaa578c94e.