

Ruby master - Bug #3217

Regexp fails to match string with '<' when encoding is UTF-8

04/29/2010 03:49 AM - brixen (Brian Shirai)

Status: Rejected	
Priority: Normal	
Assignee: naruse (Yui NARUSE)	
Target version: 1.9.2	
ruby -v: ruby 1.9.2dev (2010-04-28 trunk 27536) [i386-darwin9.8.0]	Backport:
Description =begin Hi, There is an issue matching a string like "a b c d<" when the encoding of the file is set to UTF-8 and the regexp is attempting to match 'something'. Afaik, *< is not special in the encoding. This gist illustrates the issue: http://gist.github.com/382510 Thanks, Brian =end	
Related issues: Has duplicate Backport191 - Bug #3386: Inconsistent regexp punct class matchi... Rejected 06/04/2010	

History

#1 - 05/03/2010 04:20 PM - naruse (Yui NARUSE)

- Status changed from Open to Rejected

=begin
'<' is not Punctuation on Unicode; it is Math_Symbol.
<http://unicode.org/Public/UNIDATA/extracted/DerivedGeneralCategory.txt>
=end

#2 - 05/03/2010 04:32 PM - naruse (Yui NARUSE)

- Status changed from Rejected to Assigned

- Assignee set to naruse (Yui NARUSE)

=begin
Oops, I missed this. I'll fix.
=end

#3 - 05/04/2010 05:02 AM - naruse (Yui NARUSE)

- Category set to M17N

- Status changed from Assigned to Rejected

=begin
This is feature change on Ruby 1.9.
<http://www.unicode.org/reports/tr18/>

And redcloth3's exapmle is a bug, they should use their PUNCT constant.
=end