

Redmine - Defect #40020

ScmData.binary? incorrectly considers UTF-8 text as binary

2024-01-08 15:57 - Go MAEDA

Status:	Closed	Start date:	
Priority:	Normal	Due date:	
Assignee:	Go MAEDA	% Done:	0%
Category:	SCM	Estimated time:	0.00 hour
Target version:	6.0.0	Affected version:	
Resolution:	Fixed		
Description			
Currently, the binary? method in Redmine::Scm::Adapters::ScmData often misclassifies Unicode text as binary. This is because the method actually checks whether the given data is ASCII text or not.			
The new implementation in the attached patch checks for control characters excluding tabs, newlines, and carriage returns, and calculates their proportion in the data. It ensures accurate detection of binary data while properly handling Unicode text.			

Associated revisions

Revision 22664 - 2024-01-30 02:41 - Go MAEDA

Fix ScmData.binary? method not to consider UTF-8 text as binary (#40020).

Patch by Go MAEDA (@maeda).

History

#1 - 2024-01-09 05:58 - Go MAEDA

- File 0001-Fix-ScmData.binary-method-not-to-consider-UTF-8-text.patch added

I have lowered the percentage of control characters in the data that the method considers to be binary from 0.3 to 0.1.

Since the number of control characters is approximately 11% of the 256 values that can be expressed in a single byte, I believe that a threshold value of 0.1 is enough.

#2 - 2024-01-18 08:41 - Go MAEDA

- Target version set to Candidate for next major release

#3 - 2024-01-19 09:01 - Go MAEDA

- Target version changed from Candidate for next major release to 6.0.0

Setting the target version to 6.0.0.

#4 - 2024-01-30 02:42 - Go MAEDA

- Subject changed from ScmData.binary? incorrectly identifies UTF-8 data as binary data to ScmData.binary? incorrectly considers UTF-8 text as binary

- Status changed from New to Closed

- Assignee set to Go MAEDA

- Resolution set to Fixed

Committed the fix in [r22664](#).

Files

0001-Fix-ScmData.binary-method-not-to-consider-UTF-8-text.patch	2.96 KB	2024-01-08	Go MAEDA
0001-Fix-ScmData.binary-method-not-to-consider-UTF-8-text.patch	2.96 KB	2024-01-09	Go MAEDA