Redmine - Defect #21141

Git tags and branches need to be UTF-8 encoded

2015-11-03 20:00 - Holger Just

Status:	Closed		**Start date:**	
Priority:	Normal		**Due date:**	
Assignee:			**% Done:**	0%
Category:	SCM		**Estimated time:**	0.00 hour
Target version:				
Resolution:	Duplicate		**Affected version:**	

Description

When shelling out and getting the list of git branches and tags to display in the repository navigation, the names are read in ASCII-8BIT encoding, i.e. binary strings without any encoding. This is basically correct as git itself doesn't enforce any kind of encoding for these things.

However, when rendering the navigation with branches or tags containing non-ASCII characters, ERB throws the following error in source:trunk/app/views/repositories/show.html.erb#L21

`Encoding::CompatibilityError: incompatible character encodings: ASCII-8BIT and UTF-8`

The reason for this exception is that in source:trunk/app/views/repositories/_navigation.html.erb#L18, the select box for branches is rendered. If a branch name (in ASCII-8BIT encoding) can't be converted to UTF-8 implicitly, the whole output_buffer containing the rendered HTML will be set to ASCII-8BIT encoding. When the buffer is then appended to the outer UTF-8 buffer, the error is thrown.

The solution is to force UTF-8 encoding to branch names, assuming they are stored in some UTF-8 like encoding. As we can't know the *actual* encoding of the branches, this is probably a good enough bet. In any case, it avoids HTTP 500 errors and renders something, even if the encoding doesn't fully match...

The attached patch provides the forced encoding. Unfortunately, we can't use scm_iconv with the branches as it returns a new string which would not be a Branch instance anymore.

The patch was extracted from and developed by Planio. The patch is created against the current trunk at r14793. The patched issue is present since about forever but only manifests itself only on Ruby >=1.9 and probably some kind-of-recent Rails, at least since Redmine 2.6.

Related issues:

Is duplicate of Redmine - Defect #16881: Git: repository page crashes when no...	**Closed**

History

#1 - 2015-11-03 20:12 - Toshi MARUYAMA

- Tracker changed from Patch to Defect

- Status changed from New to Closed

- Resolution set to Duplicate

Duplicate of #16881.

Git tag and branch are not always UTF-8.
Redmine test git repository filename is latin-1.
In this case tag and branch are latin-1.

#2 - 2015-11-03 20:15 - Toshi MARUYAMA

- Is duplicate of Defect #16881: Git: repository page crashes when non-ascii character in tag or branch name added

Files

0001-Force-encoding-of-git-branches-and-tags-to-UTF-8.patch	1.29 KB	2015-11-03	Holger Just