

Redmine - Feature #3396

Git: use --encoding=UTF-8 in "git log"

2009-05-20 15:00 - Vitaliy Ischenko

Status:	Closed	Start date:	2009-05-20
Priority:	Normal	Due date:	
Assignee:	Toshi MARUYAMA	% Done:	0%
Category:	SCM	Estimated time:	0.00 hour
Target version:	1.2.0		
Resolution:	Fixed		
Description			
Global setting for repositories log encoding is useless for git git has config option <code>log.outputencoding</code> if it is empty, then log encoding is UTF-8 otherwise use value specified by option			
Related issues:			
Related to Redmine - Defect # 3196: Don't properly support encoding of reposi...		Closed	2009-04-17
Related to Redmine - Feature # 1735: Per project repository log encoding setting		Closed	2008-08-03
Related to Redmine - Defect # 5251: Git: Repository path encoding of non UTF...		Closed	2010-04-07
Related to Redmine - Defect # 2664: Mercurial: Repository path encoding of no...		Closed	2009-02-04
Related to Redmine - Defect # 4773: Redmine+Git+PostgresSQL 8.4 fails with li...		Closed	2010-02-09
Related to Redmine - Defect # 7597: Subversion and Mercurial log have the pos...		Closed	2011-02-10

Associated revisions

Revision 4805 - 2011-02-11 16:30 - Toshi MARUYAMA

scm: git: prepare version string unit lib test and git log encoding (#3396).

This file includes UTF-8 literal.

We need to consider Ruby 1.9 compatibility.

Revision 4918 - 2011-02-22 01:40 - Toshi MARUYAMA

scm: git: Ruby 1.9 compatibility of adapter test (#3396).

Revision 4956 - 2011-02-27 14:49 - Toshi MARUYAMA

scm: git: add utf-8 log test in app unit test (#3396).

Revision 4959 - 2011-02-28 05:46 - Toshi MARUYAMA

scm: git: move saving changesets from adapter to model (#3396).

Revision 4961 - 2011-02-28 11:03 - Toshi MARUYAMA

scm: refactor scm log encoding test (#1735, #3396, #7597).

Bazaar log depends on locale.

On Japanese Windows, standard out is CP932.

Revision 4964 - 2011-02-28 15:12 - Toshi MARUYAMA

scm: git: use --encoding=UTF-8 in "git log" (#3396).

History

#1 - 2009-05-25 19:13 - Jean-Philippe Lang

This is pretty vague. What do you expect exactly?
I'm not a git user, so any detail is welcome.
Thanks.

#2 - 2009-05-26 10:28 - Vitaliy Ischenko

There is config option `i18n.logOutputEncoding` (per repository) in git which stores encoding for log output with git-log.

From <http://www.kernel.org/pub/software/scm/git/docs/git-config.html>

`i18n.logOutputEncoding`

Character encoding the commit messages are converted to when running git-log and friends.

if it is empty or unset, then output will be UTF-8 encoded
else value specified in this option will be used

you can get this value with ``git config i18n.logOutputEncoding``

#3 - 2009-05-29 08:35 - Jean-Philippe Lang

- *Tracker changed from Defect to Feature*

#4 - 2011-02-04 16:57 - Toshi MARUYAMA

- *Status changed from New to 7*

- *Assignee set to Toshi MARUYAMA*

#5 - 2011-02-04 17:14 - Toshi MARUYAMA

Additional reference.

<http://www.kernel.org/pub/software/scm/git/docs/git-commit-tree.html>

#6 - 2011-02-09 16:13 - Toshi MARUYAMA

Additional reference.

<http://www.kernel.org/pub/software/scm/git/docs/git.html>

| `-c <name>=<value>`

Pass a configuration parameter to the command. The value given will override values from configuration files. The <name> is expected in the same format as listed by git config (subkeys separated by dots).

#7 - 2011-02-25 16:12 - Jean-François Dagenais

I wrote [an answer](#) to Weverton Morais about how I patched a problem we had i beleive is related to this ticket. I maintain a modified linux kernel git repo, so lots of international names in there, I narrowed it down to a simple duplicating scenario.

Try making a dummy git commit with this name:

```
git commit -am"dummy test character encoding" --allow-empty --author="blažbla <tata@toto.com>"
```

Then do the changeset fetch, I use

```
ruby script/runner "Repository.fetch_changesets"
```

or the /sys/fetch_changesets with the key.

The logs will show a collation error on a query. We use git under linux platforms and never worried about encoding, so I believe our platforms default to utf8.

As my answer said, the problem seemed to be that all of the tables created by redmine (or TurnKey Linux? the base of our install.) were defaulted to latin1. In any case, the fetch_chagesets code should account for the difference in encoding if needed.

#8 - 2011-02-25 16:15 - Jean-François Dagenais

... so the point is, it's not just the file paths inside the repo, or the commit logs, but all text contained within the repo it seems.

#9 - 2011-02-28 11:00 - Vitaliy Ischenko

Jean-François Dagenais wrote:

... so the point is, it's not just the file paths inside the repo, or the commit logs, but all text contained within the repo it seems.

According to docs this is false: i18n.commitencoding relates only to *log message*, all other parts should be treated as uninterpreted sequences of non-NUL bytes (file paths, author, commiter and other commit object headers).

#10 - 2011-02-28 15:18 - Toshi MARUYAMA

- Subject changed from read Git log encoding from i18n.logoutputencoding to Git: use --encoding=UTF-8 in "git log"

#11 - 2011-02-28 15:19 - Toshi MARUYAMA

- Status changed from 7 to Closed
- Target version set to 1.2.0
- Resolution set to Fixed

Implemented until r4964.