

Redmine - Defect #5251

Git: Repository path encoding of non UTF-8 characters

2010-04-07 00:28 - Markus Mälkönen

Status:	Closed	Start date:	2010-04-07
Priority:	Low	Due date:	
Assignee:	Toshi MARUYAMA	% Done:	100%
Category:	SCM	Estimated time:	0.00 hour
Target version:	1.2.0	Affected version:	0.9.3
Resolution:	Fixed		
Description			
If filename include Scandinavian character, Redmine convert them another character. In this case, filename TEKIJÄT replaces string "TEKIJ\303\204T".			
Related issues:			
Related to Redmine - Defect # 2664: Mercurial: Repository path encoding of no...		Closed	2009-02-04
Related to Redmine - Feature # 3396: Git: use --encoding=UTF-8 in "git log"		Closed	2009-05-20
Duplicated by Redmine - Defect # 9107: Git: Redmine can't show Simplified Chi...		Closed	2011-08-23

Associated revisions

Revision 5023 - 2011-03-07 13:41 - Toshi MARUYAMA

scm: git: add core.quotePath = true in test repository config (#5251).

Revision 5024 - 2011-03-07 13:42 - Toshi MARUYAMA

scm: git: add core.quotePath = false to run git command (#5251).

Revision 5025 - 2011-03-07 14:14 - Toshi MARUYAMA

scm: git: change core.quotePath = true temporarily to run git command (#5251).

Revision 5026 - 2011-03-07 14:15 - Toshi MARUYAMA

scm: git: update test repository for path encoding (#5251).

Mercurial and Git treats file names as byte string.

This git test repository contains Latin-1 encoding path.

Be careful on non Latin-1(CP1252) Windows.

Please see r4996 comment.

Revision 5027 - 2011-03-07 14:46 - Toshi MARUYAMA

scm: git: backout r5026 (#5251).

In case git repository contains latin-1 path,

although Redmine uses "git log -C core.quotePath=false --encoding=UTF-8",

log encoding is latin-1.

Revision 5028 - 2011-03-07 14:46 - Toshi MARUYAMA

scm: git: use core.quotePath = true to run git command for database safety (#5251).

Revision 5029 - 2011-03-07 15:09 - Toshi MARUYAMA

scm: git: remove "core.quotePath = true" to run git command (#5251).

If path encoding is UTF-8, git adapter may run on Linux.

Revision 5033 - 2011-03-08 00:47 - Toshi MARUYAMA

scm: git: update test repository for path encoding (#5251).

Mercurial and Git treats file names as byte string.

This git test repository contains Latin-1 encoding path.

Be careful on non Latin-1(CP1252) Windows.

Please see r4996 comment.

I removed a revision including "copied file" from r5026 test repository.

Mercurial supports "copy", but Git does not support.

Revision 5035 - 2011-03-08 04:29 - Toshi MARUYAMA

scm: git: add instance variable for path encoding in adapter (#5251).

Revision 5036 - 2011-03-08 04:29 - Toshi MARUYAMA

scm: git: convert path encoding in "git log" (#5251).

Revision 5038 - 2011-03-08 05:50 - Toshi MARUYAMA

scm: git: support path encoding in adapter revisions() (#5251).

Revision 5039 - 2011-03-08 05:50 - Toshi MARUYAMA

scm: git: support path encoding in adapter diff (#5251).

Revision 5040 - 2011-03-08 05:50 - Toshi MARUYAMA

scm: git: support path encoding in adapter entries() (#5251).

Revision 5041 - 2011-03-08 06:17 - Toshi MARUYAMA

scm: git: support path encoding in adapter blame (#5251).

Revision 5042 - 2011-03-08 06:17 - Toshi MARUYAMA

scm: git: support path encoding in adapter cat (#5251).

Revision 5044 - 2011-03-08 07:10 - Toshi MARUYAMA

scm: git: add tests for path encoding cat, diff and blame in unit adapter test (#5251).

Revision 5049 - 2011-03-08 07:38 - Toshi MARUYAMA

scm: git: add core.quotePath = false to run git command (#5251).

Revision 5050 - 2011-03-08 07:38 - Toshi MARUYAMA

scm: git: add tests for path encoding entries() in unit adapter test (#5251).

Revision 5051 - 2011-03-08 07:49 - Toshi MARUYAMA

scm: git: prepare path encoding test in unit model test (#5251).

Revision 5057 - 2011-03-08 08:41 - Toshi MARUYAMA

scm: git: prepare path encoding test in unit model test (#5251).

Revision 5058 - 2011-03-08 08:41 - Toshi MARUYAMA

scm: git: add latest changesets path encoding test in unit model test (#5251).

Revision 5060 - 2011-03-08 12:01 - Toshi MARUYAMA

scm: git: add latin-1 encoding directory to test repository (#5251).

Revision 5061 - 2011-03-08 12:13 - Toshi MARUYAMA

scm: git: fix latin-1 directory entries() in adapter (#5251).

Revision 5062 - 2011-03-08 13:08 - Toshi MARUYAMA

scm: git: add latin-1 encoding directory test in unit adapter test (#5251).

Revision 5063 - 2011-03-08 13:08 - Toshi MARUYAMA

scm: git: add latin-1 encoding directory test in unit model test (#5251).

Revision 5064 - 2011-03-08 13:08 - Toshi MARUYAMA

scm: git: add path encoding select box at setting (#5251).

Revision 5065 - 2011-03-08 13:45 - Toshi MARUYAMA

scm: git: fix unit adapter test fails in Ruby 1.9 Linux latin-1 locale (#5251).

Revision 5066 - 2011-03-08 14:22 - Toshi MARUYAMA

scm: git: fix unit adapter test fails in Ruby 1.9 Linux latin-1 locale (#5251).

Revision 5068 - 2011-03-09 00:18 - Toshi MARUYAMA

scm: git: change core.quote_path to false in test repository config (#5251).

The -c option was introduced in git version (1.7.2)

<http://www.kernel.org/pub/software/scm/git/docs/RelNotes-1.7.2.txt>

Revision 5069 - 2011-03-09 00:30 - Toshi MARUYAMA

scm: git: switch "-c core.quote_path=false" in git version above 1.7.2 or not (#5251).

The -c option was introduced in git version (1.7.2)

<http://www.kernel.org/pub/software/scm/git/docs/RelNotes-1.7.2.txt>

Revision 5070 - 2011-03-09 02:25 - Toshi MARUYAMA

scm: git: fix PostgreSQL functional test fails (#5251).

Revision 5071 - 2011-03-09 05:13 - Toshi MARUYAMA

scm: git: unit adapter latin-1 path encoding test passes on Japanese Windows (#5251).

Ruby uses ANSI api to fork a process on Windows.

Japanese Shift_JIS and Traditional Chinese Big5 have 0x5c(backslash) problem and these are incompatible with ASCII.

Revision 5072 - 2011-03-09 05:14 - Toshi MARUYAMA

scm: git: unit model latin-1 path encoding test passes on Japanese Windows (#5251).

Ruby uses ANSI api to fork a process on Windows.

Japanese Shift_JIS and Traditional Chinese Big5 have 0x5c(backslash) problem and these are incompatible with ASCII.

Revision 5520 - 2011-04-23 05:35 - Toshi MARUYAMA

scm: use i18n string at path encoding setting (#2274, #2664, #3462, #5251).

Revision 5521 - 2011-04-23 05:35 - Toshi MARUYAMA

scm: add Japanese translation "field_scm_path_encoding" (#2274, #2664, #3462, #5251).

Revision 5522 - 2011-04-23 05:36 - Toshi MARUYAMA

scm: update locales for path encoding setting (#2274, #2664, #3462, #5251).

Revision 5523 - 2011-04-23 05:36 - Toshi MARUYAMA

scm: use i18n string at path encoding setting note (#2274, #2664, #3462, #5251).

Revision 5524 - 2011-04-23 05:37 - Toshi MARUYAMA

scm: add Japanese string at path encoding setting note (#2274, #2664, #3462, #5251).

Revision 5525 - 2011-04-23 05:37 - Toshi MARUYAMA

scm: update locales for path encoding setting note (#2274, #2664, #3462, #5251).

Revision 5628 - 2011-05-04 07:01 - Toshi MARUYAMA

scm: git: fix loss non ASCII paths if path_encoding is " (#5251).

Revision 5863 - 2011-05-21 04:16 - Toshi MARUYAMA

scm: add "path_encoding" method in abstract adapter (#2274, #3462, #2664, #5251).

Revision 5870 - 2011-05-21 04:39 - Toshi MARUYAMA

scm: git: override "path_encoding" method in adapter (#5251).

Revision 5871 - 2011-05-21 04:40 - Toshi MARUYAMA

scm: git: add unit adapter test of default path_encoding is UTF-8 (#5251).

Revision 6003 - 2011-06-08 12:05 - Toshi MARUYAMA

scm: git: skip non UTF-8 path encoding test of functional test in JRuby (#5251).

Git, Mercurial and CVS path encodings are binary.
Subversion supports URL encoding for path.
Redmine Mercurial adapter and extension use URL encoding.
Git accepts only binary path in command line parameter.
So, there is no way to use binary command line parameter in JRuby.

Revision 6004 - 2011-06-08 12:05 - Toshi MARUYAMA

scm: git: skip non UTF-8 path encoding test of unit adapter test in JRuby (#5251).

Git, Mercurial and CVS path encodings are binary.
Subversion supports URL encoding for path.
Redmine Mercurial adapter and extension use URL encoding.
Git accepts only binary path in command line parameter.
So, there is no way to use binary command line parameter in JRuby.

Revision 6005 - 2011-06-08 12:06 - Toshi MARUYAMA

scm: git: skip non UTF-8 path encoding test of unit model test in JRuby (#5251).

Git, Mercurial and CVS path encodings are binary.
Subversion supports URL encoding for path.
Redmine Mercurial adapter and extension use URL encoding.
Git accepts only binary path in command line parameter.
So, there is no way to use binary command line parameter in JRuby.

History

#1 - 2010-06-07 13:56 - Toshi MARUYAMA

Try "git config --global core.quotePath false".

#2 - 2010-06-07 15:03 - Toshi MARUYAMA

If your filename encoding is not UTF-8, try #2664 patches.

#3 - 2010-06-07 17:00 - Felix Schäfer

I'd have told you to try to set `[[RedmineSettings#Repositories-encodings]]` to whatever locale "scandinavian" is, but it seems it pertains only to file contents, not file names.

#4 - 2010-06-08 00:40 - Toshi MARUYAMA

- File *git-filename-encoding-setting.png* added

<http://www.redmine.org/issues/2664#note-4>

*Mercurial (and also Git) treats file names as byte string.
Here we need to convert them to UTF-8,
but, there's no reliable info about file name encoding.*

#5 - 2011-02-07 10:36 - Toshi MARUYAMA

- Subject changed from *Redmine can't show Scandinavian character in filenames in git repository* to *Git: Redmine can't show Scandinavian character in filenames in git repository*

#6 - 2011-02-07 10:45 - Toshi MARUYAMA

I update a patch for Git at [note-19](#) of #2664.

#7 - 2011-02-21 01:44 - Toshi MARUYAMA

- Priority changed from *Normal* to *Low*

Ruby 1.9 compatibility is very serious.

#8 - 2011-03-07 13:32 - Toshi MARUYAMA

- Subject changed from *Git: Redmine can't show Scandinavian character in filenames in git repository* to *Git: Repository path encoding of non UTF-8 characters*

- Assignee set to *Toshi MARUYAMA*

#9 - 2011-03-07 14:48 - Toshi MARUYAMA

- Status changed from *New* to *Closed*

- Assignee deleted (*Toshi MARUYAMA*)

- Resolution set to *Wont fix*

It is impossible to fix this issue in current git adapter scheme.

Please see r5027 comment.

#10 - 2011-03-07 23:58 - Toshi MARUYAMA

- Status changed from *Closed* to *Reopened*

- Assignee set to *Toshi MARUYAMA*

- Resolution deleted (*Wont fix*)

#11 - 2011-03-08 00:03 - Toshi MARUYAMA

Toshi MARUYAMA wrote:

It is impossible to fix this issue in current git adapter scheme.

Please see r5027 comment.

```
$ git cat-file commit f85f88f507577dd2fa197db9e330875b2ea0757f
tree 887f4cf35a6ed3acd3bf72843b65884d8f637029
parent 57ca437c0acbbcb749821fdf3726a1367056d364
author jsmith <jsmith@foo.bar> 1285909440 -0500
committer jsmith <jsmith@foo.bar> 1285909440 -0500
```

copy latin-1 file.

--HG--

rename : latin-1-dir/test-?.txt => latin-1-dir/test-?-1.txt

```
$ git cat-file commit 67e7792ce20ccae2e4bb73eed09bb397819c8834 | iconv -f ISO-8859-1 -t UTF-8
tree 1ec7b464f0331a7d597ee2461ec58b3a0af11114
parent 7234cb2750b63f47bf735edc50a1c0a433c2518
author test latin-1 ÀÈÁÉÂÊÃÄÅÆÇ <test@example.com> 1285909200 -0500
committer test latin-1 ÀÈÁÉÂÊÃÄÅÆÇ <test@example.com> 1285909200 -0500
encoding ISO-8859-1
```

latin-1 ÀÈÁÉÂÊÃÄÅÆÇ

Git log is binary, and f85f88f507577d log is broken.

#12 - 2011-03-08 15:53 - Toshi MARUYAMA

- % Done changed from 0 to 90

#13 - 2011-03-08 21:21 - Jean-Philippe Lang

r5049 broke all git tests with the git version currently used on the CI server (<http://www.redmine.org/builds/index.html>). The -c option was introduced in a recent git version (1.7.2). Is there any workaround for older versions?

#14 - 2011-03-09 05:22 - Toshi MARUYAMA

- Status changed from Reopened to 7

- Target version set to 1.2.0

#15 - 2011-03-09 10:47 - Toshi MARUYAMA

Jean-Philippe Lang wrote:

r5049 broke all git tests with the git version currently used on the CI server (<http://www.redmine.org/builds/index.html>). The -c option was introduced in a recent git version (1.7.2). Is there any workaround for older versions?

I fixed in r5068, r5069 and r5070.

#16 - 2011-03-09 10:50 - Toshi MARUYAMA

- *File Shift_JIS.png added*
- *Status changed from 7 to Closed*
- *% Done changed from 90 to 100*
- *Resolution set to Fixed*

I finished implementing in r5072.

Limitation

- Subversion supports URL encoding path and Redmine uses it.
Mercurial adapter helper extension wraps path with URL encoding.
Git adapter uses byte string to call "git" command.
Because Ruby uses ANSI api to fork a process on Windows,
it may have trouble in different locale on Windows.
- Git adapter uses "-c core.quotePath=false" if git version is above 1.7.2
at source:trunk/lib/redmine/scm/adapters/git_adapter.rb@5069#L342 .
If your git is older than 1.7.2, you need to set "core.quotePath=false" in "config".
- Japanese Shift_JIS and Traditional Chinese Big5 have 0x5c(backslash) problem.

Shift_JIS.png

#17 - 2011-03-29 13:42 - Toshi MARUYAMA

Toshi MARUYAMA wrote:

- *Subversion supports URL encoding path and Redmine uses it.*
Mercurial adapter helper extension wraps path with URL encoding.
Git adapter uses byte string to call "git" command.
Because Ruby uses ANSI api to fork a process on Windows,
it may have trouble in different locale on Windows.

Related Ruby issue:

Bug 1771

system()/popen()/popen3() & windows & unicode is not working

<http://redmine.ruby-lang.org/issues/show/1771>

Files

git-filename-encoding-setting.png	29.9 KB	2010-06-07	Toshi MARUYAMA
Shift_JIS.png	52.9 KB	2011-03-09	Toshi MARUYAMA