

Redmine - Defect #7064

Mercurial adapter does not recognize non alphabetic nor numeric in UTF-8 copied files

2010-12-07 13:47 - Toshi MARUYAMA

Status:	Closed	Start date:	2010-12-07
Priority:	Normal	Due date:	
Assignee:	Toshi MARUYAMA	% Done:	0%
Category:	SCM	Estimated time:	0.00 hour
Target version:	1.1.0	Affected version:	1.0.4
Resolution:	Fixed		
Description			
Before applied patch			
This is the original redmine image.			
before.png			
After applied patch			
This is the image applied patch hg-urlescape.diff .			
after.png			
Related issues:			
Related to Redmine - Defect #2664: Mercurial: Repository path encoding of non...		Closed	2009-02-04

Associated revisions

Revision 4633 - 2011-01-05 16:02 - Toshi MARUYAMA

scm: mercurial: escape any filenames by urlescape (#2664, #7064).

In Mercurial, filename is not UTF-8 but byte-string.
So it should be escaped to UTF-8-safe string before parsing as XML document.

NOTE: Mercurial's `{|obfuscate}` filter cannot be used here, because it treats filename as encoded by HGENCODING.

Contributed by Yuya Nishihara.

Revision 4634 - 2011-01-05 16:02 - Toshi MARUYAMA

scm: mercurial: update test repository for following issues and fix tests.

- copied file (#7064)
- branch, tag (#1981)

Revision 4636 - 2011-01-05 16:03 - Toshi MARUYAMA

scm: mercurial: add unit test for copied file (#7064).

History

#1 - 2010-12-07 16:42 - Yuya Nishihara

Hi, I like your patch in principle, and it should be `{name|escape}` for consistency.

But one thing, filenames in Mercurial are bytes, thus `{name|escape}` can contain invalid UTF-8 sequence for non-ASCII & non-UTF-8 filenames. It may confuse REXML parser.

The following entries have the same issue: `file`, `file_mod`, `file_add`, `file_del` and `source` .

See also [#2664](#)

#2 - 2010-12-10 09:17 - Toshi MARUYAMA

- File `settings.png` added

- File `hg-2010-12-10.diff` added

- File *mercurial_repository.tar.gz* added
- File *annotate.png* added
- File *browse.png* added
- File *copied.png* added

I merged [issue-2664-0.9-stable-2010-04-11_patch](#) of [#2664](#) and [latest-changesets-improvements-4.diff](#) of [#6159](#) excluding Subversion.
And I updated Mercurial test repository and add unit tests and functional tests.
And I attach images.

This patch fixes [#2664](#), [#5408](#), [#3449](#) and [#3567](#).

#3 - 2010-12-10 09:17 - Toshi MARUYAMA

- % Done changed from 50 to 100

#4 - 2010-12-23 15:14 - Toshi MARUYAMA

- File *japanese-windows-repo-annotate.png* added
- File *japanese-windows-repo-browse.png* added
- File *japanese-windows-repo-setting.png* added

I confirmed unit and functional tests pass for latin-1 encoding test repository on my Japanese(CP932) Windows Vista.

I post some images.

```
$ hg --version
Mercurial Distributed SCM (version 1.7.2)

$ RAILS_ENV=test ruby script/about
r:/Ruby186/lib/ruby/gems/1.8/gems/rails-2.3.5/lib/rails/gem_dependency.rb:119:Warning: Gem::Dependency#version
_requirements is deprecated and will be removed on or after August 2010. Use #requirement

About your application's environment
Ruby version      1.8.6 (i386-mingw32)
RubyGems version  1.3.7
Rack version      1.0
Rails version     2.3.5
Active Record version 2.3.5
Active Resource version 2.3.5
Action Mailer version 2.3.5
Active Support version 2.3.5
Application root  R:/work/hg-workdir/redmine-trunk-my-work
Environment       test
Database adapter  sqlite3
Database schema version 20101115000000
Loaded suite script/about
Started

Finished in 0.0 seconds.

0 tests, 0 assertions, 0 failures, 0 errors
```

#5 - 2011-01-05 11:30 - Toshi MARUYAMA

- Assignee set to Toshi MARUYAMA
- % Done changed from 100 to 0

#6 - 2011-01-06 12:58 - Toshi MARUYAMA

- Subject changed from *Mercurial adapter does not recognize non ASCII copied files* to *Mercurial adapter does not recognize non alphabetic nor numeric copied files*

#7 - 2011-01-06 13:13 - Toshi MARUYAMA

- Subject changed from *Mercurial adapter does not recognize non alphabetic nor numeric copied files* to *Mercurial adapter does not recognize non alphabetic nor numeric in UTF-8 copied files*

#8 - 2011-01-06 13:16 - Toshi MARUYAMA

- File *r4632-utf-8.png* added

- File r4633-utf-8.png added
- Status changed from New to Closed
- Target version set to 1.1.0
- Resolution set to Fixed

I confirmed to fix copying UTF-8 named file.
 I attach images.
 Non UTF-8 named file is [#2664](#) issue.
 So, I close this issue.

Files

before.png	36 KB	2010-12-07	Toshi MARUYAMA
after.png	37.2 KB	2010-12-07	Toshi MARUYAMA
hg-urlescape.diff	1.65 KB	2010-12-07	Toshi MARUYAMA
settings.png	47.6 KB	2010-12-10	Toshi MARUYAMA
hg-2010-12-10.diff	27.7 KB	2010-12-10	Toshi MARUYAMA
mercurial_repository.tar.gz	15.1 KB	2010-12-10	Toshi MARUYAMA
annotate.png	25.4 KB	2010-12-10	Toshi MARUYAMA
browse.png	114 KB	2010-12-10	Toshi MARUYAMA
copied.png	53.6 KB	2010-12-10	Toshi MARUYAMA
japanese-windows-repo-annotate.png	5.57 KB	2010-12-23	Toshi MARUYAMA
japanese-windows-repo-browse.png	18.2 KB	2010-12-23	Toshi MARUYAMA
japanese-windows-repo-setting.png	14.1 KB	2010-12-23	Toshi MARUYAMA
r4632-utf-8.png	25.7 KB	2011-01-06	Toshi MARUYAMA
r4633-utf-8.png	34.6 KB	2011-01-06	Toshi MARUYAMA