redmine.org copy 202205 - Vote #64587

Mercurial: Repository path encoding of non UTF-8 characters

2022/05/09 13:50 - Admin Redmine

:	Closed		:	2009/02/04	
:			:		
:			:	100%	
:	SCM_3		:	0.00	
:	1.2.0_27		:	0.00	
Redmine.org_URL:	https://www.redmine.org/issues/2664		**status_id:**	5	
category_id:	3		**tracker_id:**	1	
version_id:	27		**plus1:**	0	
issue_org_id:	2664		**affected_version:**		
author_id:	3753		**closed_on:**		
assigned_to_id:	11192		**affected_version_id:**		
comments:	36				

h3. Environment

* Server OS: Debian Lenny
* Redmine: svn rev 2361 (same problem with 0.8.0)
* Ruby: 1.8.6
* RubyGems: 1.3.1
* Rails: 2.1.2
* PostgreSQL: 8.3.5
* Mercurial: 1.0.1
* System locale: en_us.UTF8
* Database encoding: utf8
* Database locale: fr_FR.UTF8 (same problem with en_us.UTF8)

h3. Error

Running: @ruby script/runner "Repository.fetch_changesets" -e production@ gives the following errors:

/home/redmine/redmine-0.8.0/vendor/rails/railties/lib/commands/runner.rb:47: /home/redmine/redmine-0.8.0/vendor/rails/acti
verecord/lib/active_record/connection_adapters/abstract_adapter.rb:147:in `log':
    RuntimeError: ERROR   C22021   Minvalid byte sequence for encoding "UTF8": 0xe97365
        HThis error can also happen if the byte sequence does not match the encoding
    expected by the server, which is controlled by "client_encoding".   Fwchar.c   L1545
        Rreport_invalid_encoding: INSERT INTO "changes" ("changeset_id", "action", "revision", "branch", "from_path",
    "path", "from_revision") VALUES(781, E'A', NULL, NULL, NULL,
    E'/Quantity/doc/Présentation du projet.pdf', NULL) RETURNING "id" (ActiveRecord::StatementInvalid)
                from /home/redmine/redmine-0.8.0/vendor/rails/activerecord/lib/active_record/connection_adapters/postgresql_adapt
b:484:in `execute'
                from /home/redmine/redmine-0.8.0/vendor/rails/activerecord/lib/active_record/connection_adapters/postgresql_adapt
b:929:in `select_raw'
                from /home/redmine/redmine-0.8.0/vendor/rails/activerecord/lib/active_record/connection_adapters/postgresql_adapt
b:916:in `select'
                from /home/redmine/redmine-0.8.0/vendor/rails/activerecord/lib/active_record/connection_adapters/abstract/database
atements.rb:7:in `select_all_without_query_cache'
                from /home/redmine/redmine-0.8.0/vendor/rails/activerecord/lib/active_record/connection_adapters/abstract/query_c
e.rb:61:in `select_all'
                from /home/redmine/redmine-0.8.0/vendor/rails/activerecord/lib/active_record/connection_adapters/abstract/database
atements.rb:13:in `select_one'
                from /home/redmine/redmine-0.8.0/vendor/rails/activerecord/lib/active_record/connection_adapters/abstract/database
atements.rb:19:in `select_value'
                from /home/redmine/redmine-0.8.0/vendor/rails/activerecord/lib/active_record/connection_adapters/postgresql_adapt
b:433:in `insert'

The error seems quite similar to #834, #917, and #1663 but the error is not appening on the same table. Here, the problem comes from the "changes" table while the already reported (and corrected) issues refer a problem on the "changesets" table.

The problem seems to comes from file path which are not converted to UTF-8 (as we can notice, there is a 'é' character in the file path).

I have tried different encoding in the repository tab settings without success.

journals

I just noticed something weird with Mercurial.

When I try to remove the file mentionned above, mercurial did not success...
So the problem is maybe from Mercurial instead of Redmine.

I have the same issue with Bazaar.

I have the same issue too. My environment is a Redmine 0.8.4 in a Windows 2003 Server. My repo is Mercurial with some special character in file path, like 'ç', 'ã', 'õ'.

That's because 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.

Wei Li wrote:

> I have the same issue with Bazaar.

I'm not sure about Bazaar; but it must handle paths as UTF-8, so it seems strange.

I'm using redmine 0.9.3 on Windows Server 2003, has the same problem.

```
C:\redmine-0.9>ruby script/runner "Repository.fetch_changesets" -e production
c:/ruby/lib/ruby/gems/1.8/gems/rails-2.3.5/lib/commands/runner.rb:48: c:/ruby/li
b/ruby/gems/1.8/gems/activerecord-2.3.5/lib/active_record/connection_adapters/ab
stract_adapter.rb:219:in log': Mysql::Error: Incorrect string value: '\xB2\xE2\
xCA\xD4\xB9\xDC...' for column 'path' at row 1: INSERT INTOchanges(changese
t_id,action,revision,branch,from_path,path,from_revision) VAL
UES(279, 'A', NULL, NULL, NULL, '/doc/                        .docx', NULL) (
ActiveRecord::StatementInvalid)
        from c:/ruby/lib/ruby/gems/1.8/gems/activerecord-2.3.5/lib/active_record
/connection_adapters/mysql_adapter.rb:323:inexecute'
        from c:/ruby/lib/ruby/gems/1.8/gems/activerecord-2.3.5/lib/active_record
/connection_adapters/abstract/database_statements.rb:259:in insert_sql'
        from c:/ruby/lib/ruby/gems/1.8/gems/activerecord-2.3.5/lib/active_record
/connection_adapters/mysql_adapter.rb:333:ininsert_sql'
        from c:/ruby/lib/ruby/gems/1.8/gems/activerecord-2.3.5/lib/active_record
/connection_adapters/abstract/database_statements.rb:44:in insert_without_query
_dirty'
        from c:/ruby/lib/ruby/gems/1.8/gems/activerecord-2.3.5/lib/active_record
/connection_adapters/abstract/query_cache.rb:18:ininsert'
        from c:/ruby/lib/ruby/gems/1.8/gems/activerecord-2.3.5/lib/active_record
/base.rb:2908:in create_without_timestamps'
        from c:/ruby/lib/ruby/gems/1.8/gems/activerecord-2.3.5/lib/active_record
/timestamp.rb:53:increate_without_callbacks'
        from c:/ruby/lib/ruby/gems/1.8/gems/activerecord-2.3.5/lib/active_record
/callbacks.rb:266:in create'
        ... 30 levels...
        from c:/ruby/lib/ruby/gems/1.8/gems/rails-2.3.5/lib/commands/runner.rb:4
```

8
from c:/ruby/lib/ruby/site_ruby/1.8/rubygems/custom_require.rb:31:inge
m_original_require'
from c:/ruby/lib/ruby/site_ruby/1.8/rubygems/custom_require.rb:31:in `re
quire'

from script/runner:3

Yuya Nishihara wrote:

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

Hi, I made a patch to fix the issue.
It adds @repositories.path_encoding@ column, which can be configured via Settings -> Repository tab.

Since it changes database schema, @rake db:migrate@ is necessary. Please try it with care.

Yuya Nishihara wrote:

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

> Wei Li wrote:

>> I have the same issue with Bazaar.

> I'm not sure about Bazaar, but it must handle paths as UTF-8, so it seems strange.

I asked this Bazaar problem and #5578 at "Mercurial-ja google group":
http://groups.google.com/group/mercurial-ja/browse_thread/thread/c05134036f40207f/1d13de12bef5c7f7?#1d13de12bef5c7f7 (in Japanese).
The reason is same with #5578.
Bazaar issue: "want an option to set the output encoding, especially on win32":https://bugs.launchpad.net/bzr/+bug/340394 .
And I got a suggestion that "XMLOutput plugin":http://wiki.bazaar.canonical.com/XMLOutput is better than "bzr log".

Git problem is reported at #5251.
I tried git and Bazaar and I could display multi-bytes characters path.
This patch is for git and Bazaar.

Toshi Maruyama wrote:

> Git problem is reported at #5251.
> I tried git and Bazaar and I could display multi-bytes characters path.
> This patch is for git and Bazaar.

Git and Mercurial have absolutely the same problem, they treat filename as bytes, so the patch about Git seems reasonable.

But Bazaar's problem sounds different to me. It lies on the communication layer between Redmine and Bazaar. They should talk in UTF-8 but currently not.

To share my experence:
My system is Windows XP SP3, and Windows Server 2003.
My steps are:

1.Uninstall the redmine and reinstall it.
2.Creat hg repository in redmine folder.
3.Import the patch.
4.run "rake db:migrate RAILS_ENV=production" command

5.Restart the redmine service.

The path_encoding column was added successfully.

And I test the coding type in the list one by one, the "GBK" is correct for me.

Good luck for you!

By the way: if you have data in database, please backup it first and restore it after that the path_encoding column was added successfully.

Additionally, you need to delete repository setting created before patch applied and recreate the same repository from Redmine settings tab.

please fix this first in later version of redmine(like 1.1.2?) if #4455 Mercurial overhaul could not be done soon.
this problem stopped us from using hg for redmine completely.

These are patches for svn trunk r4799 and 1.1 stable r4800.

- attachment:20110207-impl.diff is main patch.
- attachment:20110207-db.diff is DB migration. If you have applied Yuya's attachment:issue-2664-0.9-stable-2010-04-11.patch, you don't need to apply this patch or you don't need to run "rake db:migrate". If you have not applied Yuya's patch, you need to apply this patch, and run "rake db:migrate".
- attachment:20110207-git-cvs-fs.diff is for Git, CVS and Filesystem.

This is ad hoc Mercurial adapter patch for Redmine SVN trunk and Ruby 1.9.
I confirmed to run on my Japanese Windows Vista and Mingw Ruby 1.9.2.

There is another "IO.popen" issue #6090.
source:tags/1.1.1/lib/redmine/scm/adapters/abstract_adapter.rb#L184

I think we need to refactor "IO.popen" such as "Yuya's Mercurial overhaul":
https://bitbucket.org/redmine/redmine-issue4455-snapshot/src/5b5447f4ae4b/lib/redmine/scm/adapters/mercurial_adapter.rb#cl-235

I can confirm that the patches (see note 19) solve the problem for us.
Since the issue is blocking, we would like to know if
the is a method to backout the patches and undo the schema
migration when there will be an official release that addresses this issue.

Thanks

Paolo Losi wrote:

> I can confirm that the patches (see note 19) solve the problem for us.
> Since the issue is blocking, we would like to know if
> the is a method to backout the patches and undo the schema
> migration when there will be an official release that addresses this issue.

Answering myself:

rake db:migrate:down

Sorry for the noise

_____

wow, these patches work great!!
it seems even better than before, at least now issues can be linked with r####
please make this to the next minor version 1.1.2. you have my vote. :)

there is a minor problem with the patches though. it doesn't work with codeview plugin. the error on the repository page:

NoMethodError in Code_review#update_revisions_view
Showing vendor/plugins/redmine_code_review/app/views/code_review/_update_revisions.html.erb where line #6 raised:

undefined method `review_count' for #
Extracted source (around line #6):

3: # and open the template in the editor.
4: %>
5:
6:
7: <% @changesets.each do |changeset| %>
8:     <%
9:     if changeset.review_count > 0

</pre>

<p>Toshi MARUYAMA wrote:</p>

<blockquote>
<p>These are patches for svn trunk r4799 and 1.1 stable r4800.</p>

attachment:20110207-impl.diff is main patch.
attachment:20110207-db.diff is DB migration. If you have applied Yuya's attachment:issue-2664-0.9-stable-2010-04
1.patch, you don't need to apply this patch or you don't need to run "rake db:migrate". If y
pplied Yuya's patch, you need to apply this patch, and run "rake db:migrate".
attachment:20110207-git-cvs-fs.diff is for Git, CVS and Filesystem.

</blockquote>

<hr>

<p>bo ye wrote:</p>

<blockquote>
<p>please make this to the next minor version 1.1.2. you have my vote. :)</p>
</blockquote>

<p>This feature has big behaviour change and has a db migrate.
So, I think it is difficult to apply 1.1 stable.
But, we need to consider to apply 1.2.</p>

<h2>Yuya, what do you think? </h2>

<p>Toshi MARUYAMA wrote:</p>

<blockquote>
<p>bo ye wrote:</p>

<blockquote>
<p>please make this to the next minor version 1.1.2. you have my vote. :)</p>
</blockquote>

<p>This feature has big behaviour change and has a db migrate.
So, I think it is difficult to apply 1.1 stable.
But, we need to consider to apply 1.2.</p>

<p>Yuya, what do you think?</p>

</blockquote>

<p>Same idea. For now, you can work around the issue by:</p>

<h1>put lib/redmine/scm/adapters/path_encodable_wrapper.rb</h1>

<h1>apply the patch only for app/models/repository.rb</h1>

<h1>and replace the content of @def new_scm@ method in place of @db:migrate@:</h1>

<p><pre>
scm = Redmine::Scm::Adapters::PathEncodableWrapper.new(scm, path_encoding) unless path_encoding.blank?
</pre>
   by
   <pre>
scm = Redmine::Scm::Adapters::PathEncodableWrapper.new(scm, 'encoding-name-of-your-repo')
</pre></p>

<hr>

<p>Ruby 1.9 compatibility and tests are very serious.</p>

<h2>Please see source:trunk/test/unit/lib/redmine/scm/adapters/git_adapter_test.rb@4810#L77 .</h2>

<p>Japanese Shift_JIS and Traditional Chinese Big5 have 0x5c(backslash) problem and these are incompatible with AS
Japanese EUC-JP is compatible with ASCII.</p>

<h2>Ruby uses ANSI api to fork a process on Windows.</h2>

<p>Subversion supports URL encoding for path and Redmine uses it.</p>

<h2>I think Redmine Mercurial adapter need to wrap command line path of cat, diff and annotate such as Yuya
al overhaul helper extension.</h2>

<p>I start implementing in new way.</p>

<h2>Ruby 1.9 compatibility is very serious. </h2>

<hr>

<hr>

<hr>

<p>I can't run on my Japanese Windows Ruby 1.9.2 without #4050 "Ruby-1.9-Encoding.default_external.diff&q
a href="http://www.redmine.org/attachments/5498/Ruby-1.9-Encoding.default_external.diff">http://www.redmine.org/attachment
s/5498/Ruby-1.9-Encoding.default_external.diff .
Despite applying this patch, I got following error.</p>

<pre>
[2011-03-04 20:51:58] ERROR Encoding::InvalidByteSequenceError: "\x9C" followed by "-" on Windows-31J
                r:/Ruby192/lib/ruby/gems/1.9.1/gems/rails-2.3.11/lib/rails/rack/static.rb:37:in `file?'
                r:/Ruby192/lib/ruby/gems/1.9.1/gems/rails-2.3.11/lib/rails/rack/static.rb:37:in `file_exist?'
                r:/Ruby192/lib/ruby/gems/1.9.1/gems/rails-2.3.11/lib/rails/rack/static.rb:18:in `call'
                r:/Ruby192/lib/ruby/gems/1.9.1/gems/rack-1.1.0/lib/rack/urlmap.rb:47:in `block in call'
                r:/Ruby192/lib/ruby/gems/1.9.1/gems/rack-1.1.0/lib/rack/urlmap.rb:41:in `each'
                r:/Ruby192/lib/ruby/gems/1.9.1/gems/rack-1.1.0/lib/rack/urlmap.rb:41:in `call'
                r:/Ruby192/lib/ruby/gems/1.9.1/gems/rails-2.3.11/lib/rails/rack/log_tailer.rb:17:in `call'
                r:/Ruby192/lib/ruby/gems/1.9.1/gems/rack-1.1.0/lib/rack/content_length.rb:13:in `call'
                r:/Ruby192/lib/ruby/gems/1.9.1/gems/rack-1.1.0/lib/rack/handler/webrick.rb:48:in `service'
                r:/Ruby192/lib/ruby/1.9.1/webrick/httpserver.rb:111:in `service'
                r:/Ruby192/lib/ruby/1.9.1/webrick/httpserver.rb:70:in `run'
                r:/Ruby192/lib/ruby/1.9.1/webrick/server.rb:183:in `block in start_thread'
</pre>

<hr>

<p>"Files" module has similar strange behavior on my Japanese Windows Ruby 1.9.2.</p>

<h2>I give up fix it.</h2>

<p>I finished implementing this feature until r5001.
And I confirmed to run on my Japanese Windows Ruby 1.8 and Linux Ruby 1.8.</p>

<h2>On Linux with #4050 "Ruby-1.9-Encoding.default_external.diff"http://www.redmine.org/attachments/5498/Ruby-1.9-Encoding.default_external.diff, I confirmed to run in ISO-8859-1 locale.</h2>

related_issues

relates,Closed,5251,Git: Repository path encoding of non UTF-8 characters
relates,Closed,2274,Filesystem Repository path encoding of non UTF-8 characters
relates,Closed,3462,CVS: Repository path encoding of non UTF-8 characters
relates,Closed,7064,Mercurial adapter does not recognize non alphabetic nor numeric in UTF-8 copied files
relates,New,2799,Support for Bazaar's shared reposetories (created with init-repo)
relates,Closed,6090,Most binary files become corrupted when downloading from CVS repository browser when Redmine is running on a Windows server
relates,Closed,4050,Ruby 1.9 support
relates,Closed,3396,Git: use --encoding=UTF-8 in "git log"
relates,Closed,4773,Redmine+Git+PostgresSQL 8.4 fails with linux kernel tree (encoding)
duplicates,Closed,5408,Mercurial and chinese code
duplicates,Closed,3677,fetching changesets from Mercurial repository fails
duplicates,Closed,8726,Redmine+Mercurial+PostgreSQL 9 falls with cyrrilic filenames in repository

#1 - 2022/05/10 17:26 - Admin Redmine

- SCM_3

- 1.2.0_27