

Redmine - Defect #24030

When SVN or Git repository has a commit comment include an emoji (4 bytes character), error occurs

2016-10-10 05:47 - Jiamin Wang

Status: Closed	Start date:
Priority: Normal	Due date:
Assignee:	% Done: 0%
Category: SCM	Estimated time: 0.00 hour
Target version:	Affected version: 3.1.2
Resolution: Invalid	
Description I already set the MySQL and all databases' charset to utf8mb4 the error is in fetch_changeset method, when insert the commit comment into database table	
Related issues: Related to Redmine - Feature #31921: Changes to properly support 4 byte chara... Closed	

Associated revisions

Revision 19698 - 2020-04-14 10:20 - Toshi MARUYAMA

scm: git: add new test repository which has UTF-8 Emoji committer and comments (#24030)

Revision 19699 - 2020-04-14 10:20 - Toshi MARUYAMA

scm: git: add test of UTF-8 Emoji committer and comments (#24030)

Revision 19700 - 2020-04-14 10:47 - Toshi MARUYAMA

fix source indent of previous commit (#24030)

Revision 19701 - 2020-04-14 14:04 - Toshi MARUYAMA

scm: git: replace test repository text file content meaningfully (#24030)

Revision 19706 - 2020-04-18 06:59 - Go MAEDA

Fix a RuboCop offense Style/ColonMethodCall due to r19699 (#24030).

History

#1 - 2016-10-10 05:59 - Jiamin Wang

error occurs in save_revision method in /app/models/git.rb or /app/models/subversion.rb

#2 - 2016-10-10 09:02 - Toshi MARUYAMA

- Category changed from Database to SCM

#3 - 2019-07-31 09:46 - Tobias Fischer

This is still an issue in Redmine 3.4.11

Environment:

```
Redmine version      3.4.11.stable.18254
Ruby version         2.2.1-p85 (2015-02-26) [x86_64-linux]
Rails version        4.2.11.1
Environment           production
Database adapter     Mysql2
```

SCM:

```
Subversion           1.6.17
Git                   1.7.9.5
Filesystem
```

Error message in log:

```
Started GET "/projects/PROJECT/repository/REPO_ID" for 37.24.44.210 at 2019-07-31 09:43:30 +0200
Processing by RepositoriesController#show as HTML
  Parameters: {"id"=>"PROJECT", "repository_id"=>"REPO_ID"}
  Current user: USERNAME (id=5)
Completed 500 Internal Server Error in 145ms (ActiveRecord: 13.2ms)
```

```
ActiveRecord::StatementInvalid (Mysql2::Error: Incorrect string value: '\xF0\x9F\xA4\xA6' for column 'comments'
 at row 1: INSERT INTO `changesets` (`repository_id`, `revision`, `scmid`, `committer`, `commit_date`, `commi
tted_on`, `comments`, `user_id`) VALUES (18, 'a5814b480017f7702d46d084dalaala7cb5d05b0', 'a5814b480017f7702d46
d084dalaala7cb5d05b0', 'Tobias Fischer <tobias.fischer@domain.tld>', '2019-07-04', '2019-07-04 21:35:45', '[ci
] gradle file bei vorherigem Commit vergessen
```

#4 - 2019-07-31 09:51 - Tobias Fischer

Funny thing: Also the comment here above get's truncated at the position of the emoji in my log message. Emojis are a big issue for Redmine!

#5 - 2019-08-17 18:30 - Marius BĂLTEANU

- Related to Feature #31921: Changes to properly support 4 byte characters (emoji) when database is MySQL added

#6 - 2020-04-15 18:16 - Toshi MARUYAMA

- File error.png added

- File env.png added

I have added new test repository and tests because existing test repository encoding is 'ISO-8859-1'.

source:trunk/test/unit/repository_git_test.rb@19705#L42

On my CentOS 7 MySQL 5.5.62, utf8mb4 environment passes test.

```
mysql> show variables like "character%";
+-----+-----+
| Variable_name | Value |
+-----+-----+
| character_set_client | utf8mb4 |
| character_set_connection | utf8mb4 |
| character_set_database | utf8mb4 |
| character_set_filesystem | binary |
| character_set_results | utf8mb4 |
| character_set_server | utf8mb4 |
| character_set_system | utf8 |
| character_sets_dir | /usr/share/mysql/charsets/ |
+-----+-----+
8 rows in set (0.00 sec)

mysql> show create table changesets\G
***** 1. row *****
Table: changesets
Create Table: CREATE TABLE `changesets` (
  `id` int(11) NOT NULL AUTO_INCREMENT,
  `repository_id` int(11) NOT NULL,
  `revision` varchar(255) NOT NULL,
  `committer` varchar(255) DEFAULT NULL,
  `committed_on` datetime NOT NULL,
  `comments` longtext,
  `commit_date` date DEFAULT NULL,
  `scmid` varchar(255) DEFAULT NULL,
  `user_id` int(11) DEFAULT NULL,
  PRIMARY KEY (`id`),
  UNIQUE KEY `changesets_repos_rev` (`repository_id`,`revision`),
  KEY `index_changesets_on_committed_on` (`committed_on`),
  KEY `changesets_repos_scmid` (`repository_id`,`scmid`),
  KEY `index_changesets_on_repository_id` (`repository_id`),
  KEY `index_changesets_on_user_id` (`user_id`)
) ENGINE=InnoDB AUTO_INCREMENT=114 DEFAULT CHARSET=utf8mb4 ROW_FORMAT=DYNAMIC
1 row in set (0.00 sec)
```

With this change,

```
diff --git a/test/unit/repository_git_test.rb b/test/unit/repository_git_test.rb
```

```

--- a/test/unit/repository_git_test.rb
+++ b/test/unit/repository_git_test.rb
@@ -592,7 +592,8 @@ class RepositoryGitTest < ActiveSupport:
  end

  if File.directory?(REPOSITORY_UTF8_PATH) &&
-   !(Redmine::Database::mysql? && !is_mysql_utf8mb4)
+   # !(Redmine::Database::mysql? && !is_mysql_utf8mb4)
+   true
  def test_utf8_emoji
    repo = Repository::Git.create(
      :project => @project,

```

NOT utf8mb4 environment causes error.

```

Error:
RepositoryGitTest#test_utf8_emoji:
ActiveRecord::StatementInvalid: Mysql2::Error: Incorrect string value: '\xF0\x9F\x98\x83 <...
' for column 'committer' at row 1: INSERT INTO `changesets` (`repository_id`, `revision`, `co
mmitter`, `committed_on`, `comments`, `commit_date`, `scmid`) VALUES (27, 'd37ec5b2c54b6d1b87
5888f07571e372acd638c9', 'U+1F603😄 <none@none>', '2020-04-12 23:45:24', 'U+1F603😄', '2020-0
4-12', 'd37ec5b2c54b6d1b875888f07571e372acd638c9')
app/models/repository/git.rb:223:in `save_revision'
app/models/repository/git.rb:212:in `block (2 levels) in save_revisions'
app/models/repository/git.rb:209:in `block in save_revisions'
app/models/repository/git.rb:208:in `each'
app/models/repository/git.rb:208:in `save_revisions'
app/models/repository/git.rb:158:in `fetch_changesets'
test/unit/repository_git_test.rb:606:in `test_utf8_emoji'

bin/rails test test/unit/repository_git_test.rb:597

```

#7 - 2020-04-15 19:02 - Toshi MARUYAMA

- File *subversion_repository.dump.gz* added
- File *svn.diff* added

This is Subversion test repository and test code.
Same results with git.

I give up to commit because it is very risky
if [CI server](#) is NOT utf8mb4, tests skip.

#8 - 2020-04-15 19:08 - Toshi MARUYAMA

- File *mercurial.hg* added
- File *mercurial.png* added

This is Mercurial bundle file and its results.

root

Name	Size	Revision	Age	Author	Comment
 test.txt	260 Bytes	0:acfb23e0de38	3 days	U+1F603☹️	U+1F603☹️

Latest revisions

#	Date	Author	Comment
• 0:acfb23e0de38	04/12/2020 11:45 PM	U+1F603☹️	U+1F603☹️

[View all revisions](#) | [View revisions](#)

Also available in:  Atom

```
$  
$ LANGUAGE=C hg log -l 1  
changeset: 0:acfb23e0de38  
bookmark: master  
tag: tip  
user: U+1F603☹️  
date: Sun Apr 12 23:45:24 2020 +0900  
summary: U+1F603☹️
```

#9 - 2020-04-15 19:15 - Marius BĂLTEANU

- Priority changed from High to Normal

Toshi, what do you think if we assign this to [4.2.0](#)? We already have a ticket there for utf8mb4 support on mysql ([#32054](#)).

#10 - 2020-04-15 19:16 - Toshi MARUYAMA

- Status changed from New to Closed

- Resolution set to Invalid

Conclusion:

No error with correct server configuration and "encoding: utf8mb4" of database.yml.
[source:trunk/config/database.yml.example@19705#L12](#)

#11 - 2020-04-15 19:18 - Toshi MARUYAMA

Marius BALTEANU wrote:

Toshi, what do you think if we assign this to [4.2.0](#)? We already have a ticket there for utf8mb4 support on mysql, please see [#32054](#).

Good idea.

#12 - 2020-04-20 04:33 - Toshi MARUYAMA

- File same-with-r19706.diff added

This is additional patch.
"use same "Redmine::Database.mysql?" as with [r19706](#)".

Files

error.png	34.2 KB	2020-04-15	Toshi MARUYAMA
env.png	49.1 KB	2020-04-15	Toshi MARUYAMA
subversion_repository.dump.gz	113 KB	2020-04-15	Toshi MARUYAMA
svn.diff	4.8 KB	2020-04-15	Toshi MARUYAMA
mercurial.hg	445 Bytes	2020-04-15	Toshi MARUYAMA
mercurial.png	36.8 KB	2020-04-15	Toshi MARUYAMA
same-with-r19706.diff	1.95 KB	2020-04-20	Toshi MARUYAMA