Redmine - Defect #5578

Bazaar: Missing characters from repository comments

2010-05-23 23:56 - Chris G

Status:	Closed	**Start date:**	2010-05-23
Priority:	Normal	**Due date:**	
Assignee:		**% Done:**	0%
Category:	SCM	**Estimated time:**	0.00 hour
Target version:			
Resolution:	Invalid	**Affected version:**	

Description

When viewing a project repository, the commit messages have missing characters. The repository is Bazaar and the comments have international characters that normally show up when using the Bazaar GUI but when viewed in Redmine, the special-characters are simply omitted.

Although the commit "comments" or "messages" have missing characters, when viewing the repository files using Redmine or doing "diff" on them, special-characters are displayed normally.

Under Settings -> Repository I have "Repositories encodings" set to "utf-8" and "Commit messages encoding" set to "UTF-8". It doesn't seem to differ what I set "Commit messages encoding" to, the special-characters never show up, not even incorrectly.

No errors or warnings are generated. This issue is similar to Issue #2133.

Database: MySQL 5.0.41
Ruby version: ruby 1.8.7 (2010-01-10 patchlevel 249) [i386-mingw32]
Rails version: rails 2.3.5
Redmine version: redmine 0.9.4

Example characters that go missing: áéíýúðþ (Icelandic special-characters)

Related issues:

Has duplicate Redmine - Defect #5142: Can`t import Bazaar changesets from rep...	**Closed**	**2010-03-22**
Has duplicate Redmine - Defect #8385: Character encoding is wrong in reposito...	**Closed**	**2011-05-17**

History

#1 - 2010-05-24 06:12 - Toshi MARUYAMA

Check your MySQL encoding.
Refer http://www.redmine.org/issues/4455#note-31 .

#2 - 2010-05-24 14:21 - Chris G

So I checked the table 'changesets' in the redmine database, and yes, the characters are missing. I deleted the repository (from the redmine settings) and flushed the 'changes' and 'changesets' tables. Then I did:

```
ALTER TABLE changes CONVERT TO CHARACTER SET utf8 COLLATE utf8_general_ci
ALTER TABLE changesets CONVERT TO CHARACTER SET utf8 COLLATE utf8_general_ci
ALTER TABLE comments CONVERT TO CHARACTER SET utf8 COLLATE utf8_general_ci
```

The 'changesets' table is where the commit messages are stored. I then added the repository again and still the special-characters are missing.

Mind that the Bazaar data is not stored in a DB, it's in plain text files under the .bzr folder. Upon loading the repository for the first time, the commit messages are transferred to the redmine DB and somewhere along the way the special-characters are omitted.

I probably should change the character set for the rest of the tables (which is 'latin1_general_ci') to utf8 but Peter Fern's convert command in Issue #4455 doesn't work as expected on Cyqwin.

What can I do to debug this issue? Any help is appreciated.

#3 - 2010-05-24 16:28 - Toshi MARUYAMA

Check your database "encoding" in REDMINE_DIR/config/database.yml.

#4 - 2010-05-24 17:11 - Chris G

'encoding' for production in database.yml defaults to 'utf8'

#5 - 2010-05-24 17:32 - Toshi MARUYAMA

Chris G wrote:

> What can I do to debug this issue? Any help is appreciated.

Redmine executes "bzr log" at source:tags/0.9.4/lib/redmine/scm/adapters/bazaar_adapter.rb#L80 .

#6 - 2010-05-24 18:35 - Chris G

Doing 'bzr log -v --show-ids' on the command line shows the characters correctly. If I look at ~/.bzr.log I can see that the output encoding is set to 'cp1252'

```
encoding stdout as osutils.get_user_encoding() 'cp1252'
```

This seems to be causing issues for other people as well, see https://bugs.launchpad.net/bzr/+bug/340394

There they are trying to change the command so that you can do 'bzr log --encoding_type "utf-8"' and thus get the right encoding format.

Would it be possible to have Redmine convert the encoding from cp1252 to utf8 and thus overcome the problem?

#7 - 2010-05-24 19:41 - Felix Schäfer

Chris G wrote:

> Would it be possible to have Redmine convert the encoding from cp1252 to utf8 and thus overcome the problem?

That would be the commit messages encoding setting.

#8 - 2010-05-24 22:49 - Chris G

Setting 'Commit messages encoding' in Admin -> Settings to windows-1252 and re-adding the repository fixes the problem. I'm pretty sure I tried this before, though it may have been before I converted the MySQL tables to utf8. The issue has been resolved.

#9 - 2010-05-24 23:37 - Felix Schäfer

- Status changed from New to Closed

- Resolution set to Invalid

#10 - 2010-06-05 13:28 - Toshi MARUYAMA

Bazaar issue want an option to set the output encoding, especially on win32 has a file name encoding problem too.
Now I post a message at #2664.

#11 - 2011-02-22 09:40 - Toshi MARUYAMA

- Subject changed from Missing characters from repository comments to Bazaar: Missing characters from repository comments

#12 - 2011-02-22 09:40 - Toshi MARUYAMA

- Category set to SCM