

Redmine - Defect #11089

UTF-8 encoding not showing correctly when looking highlighted php file contents

2012-06-04 19:13 - Troex Nevelin

Status:	Closed	Start date:	
Priority:	Normal	Due date:	
Assignee:		% Done:	0%
Category:	Text formatting	Estimated time:	0.00 hour
Target version:		Affected version:	2.0.1
Resolution:			
Description			
Ruby version 1.9.3 (x86_64-linux) RubyGems version 1.8.11 Rack version 1.4 Rails version 3.2.5 Database adapter mysql2 Database schema version 20120422150750 Git version 1.7.2.5			
When I request file to see it contents (repository/revisions/HASH/entry) instead UTF-8 text I get '???'. I'm using Git SCM and my files are valid UTF-8 (without BOM). I have this problem with Chineses, Russian, Thai and other scripts than latin. However seeing diff's and attached utf-8 files are okay.			
Related issues:			
Has duplicate Redmine - Defect #11131: repository View and Annotate code Utf-...			Closed
Has duplicate Redmine - Defect #14445: <code> inside <pre> destroy polish dia...			Closed

History

#1 - 2012-06-04 20:40 - Etienne Massip

Did you set any value in the Attachments and repositories encodings setting (in Administration/Settings General tab)?

If not, try to?

#2 - 2012-06-05 00:34 - Troex Nevelin

Yes I have tried setting it to UTF-8 but it has no effect

#3 - 2012-06-05 00:52 - Toshi MARUYAMA

- Subject changed from UTF-8 encoding not showing correctly when looking file contents to Git: encoding not showing correctly when looking file contents

#4 - 2012-06-05 01:03 - Toshi MARUYAMA

Redmine uses "git show".

[source:tags/2.0.1/lib/redmine/scm/adapters/git_adapter.rb#L372](https://sourceforge.net/p/redmine/scm/adapters/git_adapter.rb/L372)

Git 1.7.3.4, "git show --help" says

```
The contents of the blob objects are uninterpreted sequences of bytes.  
There is no encoding translation at the core level.
```

#5 - 2012-06-05 13:53 - Troex Nevelin

I understand that git stores files in binary form, but calling from console:

```
git show --no-color HEAD:../lang/ru/component.php
```

returns UTF-8 valid text, as I understand Redmine tries to guess encoding and sanitise content making sure no invalid characters pass to view.

For example [source:trunk/config/locales/ja.yml](https://sourceforge.net/p/redmine/scm/adapters/git_adapter.rb/L372) this displays up correctly (but it uses SVN).

I think there is encoding guess problem in [source:tags/2.0.1/lib/redmine/codeset_util.rb#L84](https://source.ubuntu.com/ubuntu/+source/redmine/commit/d7e2a66d) calling `.to_utf8_by_setting_internal(str)` sets ASCII-8BIT encoding on line 94?

#6 - 2012-06-05 15:13 - Toshi MARUYAMA

- File *gh-new-d7e2a66d.png* added

I cannot reproduce.

<https://github.com/redmine/redmine/commit/d7e2a66d>

Could you attach this "git show" output file?

#7 - 2012-06-05 15:36 - Troex Nevelin

- File *git-show-component.php.txt* added

```
git show --no-color HEAD:../lang/ru/component.php > git-show-component.php.txt
```

I'm running Redmine on Debian 6, with ruby 1.9.3p125 (2012-02-16) [x86_64-linux] package compiled from debian ruby repository, using unicorn rack server.

I'm almost sure this is local related problem. Can you guide me how to debug this problem? I'm familiar with ruby and ror. I have tried to output raw **content** in `app/views/common/_file.html.erb` but it gives me `ActionView::Template::Error (incompatible character encodings: UTF-8 and ASCII-8BIT)` error

#8 - 2012-06-05 15:39 - Troex Nevelin

- File *git-show-component.php* added

#9 - 2012-06-05 15:45 - Troex Nevelin

- File *issue-attached-files.png* added

- File *test-file-with-php-ext.png* added

- File *test-file-with-txt-ext.png* added

I've made one more test on my setup, I've attached the same file to an issue but with different extensions `.txt` and `.php` and when trying to see attached file I get an issue with viewing syntax highlighted file. So this is not only Git related problem.

But no issue here in this ticket.

```
# grep coderay Gemfile.lock
  coderay (1.0.6)
  coderay (~> 1.0.6)
```

#10 - 2012-06-05 16:19 - Toshi MARUYAMA

- Subject changed from *Git: encoding not showing correctly when looking file contents to UTF-8 encoding not showing correctly when looking file contents*

- Category deleted (SCM)

#11 - 2012-06-11 10:52 - gehao liu

- Status changed from *New* to *Resolved*

same problem!!!!!!!!!!

txt extname is OK,

python extname `.py` is OK.

php extname is wrong,

problem is viewing syntax highlighted!!!!!!!!!!

#12 - 2012-06-11 11:09 - gehao liu

- File *def.php* added

- File *def.py* added

- File *def.txt* added

gehao liu wrote:

same problem!!!!!!!!!!

txt extname is OK,

python extname .py is OK.
php extname is wrong,
problem is viewing syntax highlighted!!!!!!!!!!

#13 - 2012-06-11 11:20 - Toshi MARUYAMA

- Status changed from Resolved to New

#14 - 2012-06-12 09:53 - gehao liu

this is coderay 1.0.6's bug,only php file.

#15 - 2013-03-27 10:45 - András Kolesár

coderay php encoding issue has been solved:

<https://github.com/rubychan/coderay/issues/40>

checked, works fine with updated coderay/scanners/php.rb file

#16 - 2013-03-27 14:23 - Etienne Massip

- Subject changed from UTF-8 encoding not showing correctly when looking file contents to UTF-8 encoding not showing correctly when looking highlighted file contents

- Category set to Text formatting

- Status changed from New to Confirmed

- Target version set to Candidate for next minor release

Upgrade dep to 1.0.9 or 1.1.

#17 - 2013-06-04 15:33 - Toshi MARUYAMA

- Related to Defect #13692: warning: already initialized constant on Ruby 1.8.7 added

#18 - 2013-06-04 15:35 - Toshi MARUYAMA

- Related to deleted (Defect #13692: warning: already initialized constant on Ruby 1.8.7)

#19 - 2013-06-04 15:41 - Toshi MARUYAMA

- File def.php.png added

- Status changed from Confirmed to Closed

Coderay version is defined "~> 1.0.6" in [source:tags/2.3.1/Gemfile#L6](#)

So, Coderay 1.0.9 is installed in [2.3.1](#).

This is note 12 php image.

```
1 <?php
2 #这是中文
3 phpinfo();
```

#20 - 2013-06-04 15:41 - Toshi MARUYAMA

- Subject changed from UTF-8 encoding not showing correctly when looking highlighted file contents to UTF-8 encoding not showing correctly when looking highlighted php file contents

#21 - 2013-06-04 16:23 - Toshi MARUYAMA

- Target version deleted (Candidate for next minor release)

#22 - 2013-07-20 07:26 - Toshi MARUYAMA

- Has duplicate Defect #14445: <code> inside <pre> destroy polish diacritic of PHP added

Files

utf-8-not-shown-in-file-contents-view.png	50.4 KB	2012-06-04	Troex Nevelin
---	---------	------------	---------------

diff-view-is-okay.png	56.6 KB	2012-06-04	Troex Nevelin
gh-new-d7e2a66d.png	53.8 KB	2012-06-05	Toshi MARUYAMA
git-show-component.php.txt	2.61 KB	2012-06-05	Troex Nevelin
git-show-component.php	2.61 KB	2012-06-05	Troex Nevelin
issue-attached-files.png	15.2 KB	2012-06-05	Troex Nevelin
test-file-with-php-ext.png	145 KB	2012-06-05	Troex Nevelin
test-file-with-txt-ext.png	93.5 KB	2012-06-05	Troex Nevelin
def.php	31 Bytes	2012-06-11	gehao liu
def.py	38 Bytes	2012-06-11	gehao liu
def.txt	31 Bytes	2012-06-11	gehao liu
def.php.png	3.93 KB	2013-06-04	Toshi MARUYAMA