

redmineorg-copy202205 - Vote #78059

Should not replace all invalid utf8 characters (e.g in mail)

2022/05/09 18:09 - Admin Redmine

ステータス:	Closed	開始日:	2022/05/09
優先度:	通常	期日:	
担当者:		進捗率:	0%
カテゴリ:	I18n_37	予定工数:	0.00時間
対象バージョン:	3.4.0_119	作業時間:	0.00時間
Redmineorg_URL:	https://www.redmine.org/issues/24616	status_id:	5
category_id:	37	tracker_id:	1
version_id:	119	plus1:	0
issue_org_id:	24616	affected_version:	
author_id:	123153	closed_on:	
assigned_to_id:	0	affected_version_id:	
comments:	9		

説明

Hello,
I've an email, that is encoded in utf8, but it contains an invalid character. In this case, redmine converts the content to us-ascii and then to utf8. This step will replace non-ascii compatible chars to "?". Why?

1) Failure:

MailHandlerTest#test_invalid_utf8 [/test/unit/mail_handler_test.rb:548]:

Expected: "З д р а в с т в у й т е ?"

Actual: "?????????????"

I changed Redmine::CodesetUtil.replace_invalid_utf8(str) and Redmine::CodesetUtil.to_utf8(str, encoding)

```
str = str.encode("US-ASCII", :invalid => :replace, :undef => :replace, :replace => '?').encode("UTF-8")
```

to

```
str = str.encode("UTF-8", :invalid => :replace, :undef => :replace, :replace => '?')
```

all tests are passing with this change.

```
Redmine version      3.3.1.stable
Ruby version         2.1.5-p273 (2014-11-13) [x64-mingw32]
Rails version        4.2.7.1
Environment          production
Database adapter      Mysql2
SCM:
  Git                 2.10.1
  Filesystem
Redmine plugins:
  no plugin installed
```

journals

Looks good to me.

```
# valid UTF-8 string
text = "こんにちは"
p text.valid_encoding? # => true
```

```
# making invalid UTF-8 string
text.force_encoding('ASCII-8BIT')
text[-1] = 0xff.chr
text.force_encoding('UTF-8')
p text.valid_encoding? # => false
```

```
p text # => "こんにちは\xE3\x81\xFF"

# Current code of Redmine
p text.encode("US-ASCII", :invalid => :replace, :undef => :replace, :replace => '?').encode("UTF-8")
# => "??????"

# Fixed code by Pavel Rosický
p text.encode("UTF-8", :invalid => :replace, :undef => :replace, :replace => '?')
# => "こんにちは?"
```

----- Did you run whole tests? Especially this test.
source:tags/3.3.1/test/unit/lib/redmine/codeset_util_test.rb ----- Pavel Rosický
wrote: > Hello, > I've an email, that is encoded in utf8, but it contains an invalid character. In this case, redmine converts the content to us-ascii and then to utf8. This step will replace non-ascii compatible chars to "?". Why? You can see this function purpose.
source:tags/3.3.1/test/unit/lib/redmine/codeset_util_test.rb#L68 ----- Thanks
Toshi, I rechecked it again and all tests are passing. source:tags/3.3.1/test/unit/lib/redmine/codeset_util_test.rb#L68 In this case, my change has no effect on the result, because the string contains just one invalid utf-8 character.

```
s1.encode('us-ascii', :invalid => :replace, :undef => :replace, :replace => '?').encode('utf-8')
"Texte encod? en ISO-8859-1."

# patched
s1.encode('utf-8', :invalid => :replace, :undef => :replace, :replace => '?')
"Texte encod? en ISO-8859-1."
```

but a combination of valid and invalid utf-8 chars (non-ascii-compatible) will result both characters are stripped. Try out GO Media's example.

```
$ irb
1.9.3-p551 :001 > text = "こんにちは\xE3\x81\xFF"
=> "こんにちは\xE3\x81\xFF"
1.9.3-p551 :002 > text = text.encode("UTF-8", :invalid => :replace, :undef => :replace, :replace => '?')
=> "こんにちは\xE3\x81\xFF"
1.9.3-p551 :003 > text.valid_encoding?
=> false

$ irb
2.3.3 :001 > text = "こんにちは\xE3\x81\xFF"
=> "こんにちは\xE3\x81\xFF"
2.3.3 :002 > text = text.encode("UTF-8", :invalid => :replace, :undef => :replace, :replace => '?')
=> "こんにちは?"
2.3.3 :003 > text.valid_encoding?
=> true
```

Pavel Rosický wrote:

Hello,
I've an email, that is encoded in utf8, but it contains an invalid character. In this case, redmine converts the content to us-ascii and then to utf8. This step will replace non-ascii compatible chars to "?". Why?

Because of Ruby 1.8.7 behavior compatibility.
source:tags/2.6.9/lib/redmine/codeset_util.rb

I have committed r16273 to pass on Ruby 1.9.3.

I don't want to change behavior on stable.

履歴

#1 - 2022/05/10 17:06 - Admin Redmine

- カテゴリ を 118n_37 にセット
- 対象バージョンを 3.4.0_119 にセット