

## Redmine - Defect #2274

### Filesystem Repository path encoding of non UTF-8 characters

2008-12-04 17:31 - Toni Kerschbaum

<b>Status:</b>	Closed	<b>Start date:</b>	2008-12-04
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Toshi MARUYAMA	<b>% Done:</b>	100%
<b>Category:</b>	SCM	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	1.2.0	<b>Affected version:</b>	
<b>Resolution:</b>	Fixed		
<b>Description</b>			
<p>The filesystem repository has some issues regarding special characters in filenames or directory names. If the name of a file or directory contains a special char (like ü, ö, ä for example), the codepage will get mixed up and the browser will have trouble displaying it (see attached pictures). If one corrects the codepage manually, the files in the repo will be displayed correctly, but every other special char (table headers, etc.) will get messed up.</p> <p>In addition to that, if a directory contains a special char, it is browseable, but every subdir and file in that directory is not. That is mainly because every '/' after the special char will be converted to %2F and every special char will be converted to its entity, completely messing up the URI. You will get a 404 (not found) error. So, for instance</p> <p><a href="http://redmine.my.domain/repositories/browse/project/Brücke">http://redmine.my.domain/repositories/browse/project/Brücke</a> -&gt; <a href="http://redmine.my.domain/repositories/browse/project/Br%FCcke">http://redmine.my.domain/repositories/browse/project/Br%FCcke</a> <a href="http://redmine.my.domain/repositories/browse/project/Brücke/übersicht.jpg">http://redmine.my.domain/repositories/browse/project/Brücke/übersicht.jpg</a> -&gt; <a href="http://redmine.my.domain/repositories/browse/project/Br%FCcke%2F%FCbersicht.jpg">http://redmine.my.domain/repositories/browse/project/Br%FCcke%2F%FCbersicht.jpg</a></p> <p>Manually replacing every %2F with / did not help really. You will get the Redmine page, but an error saying that the file does not exist in the repo. Replacing every %-code with its special char will lead to a 500 (internal) error.</p> <p>Setting the codepages under Administration -&gt; Repositories -&gt; Codepages does not affect this in any way. Tried settings:</p> <ul style="list-style-type: none"><li>- UTF-8, ISO 8859-1, ISO 8859-15, CP1252</li><li>- ISO-8859-1, ISO-8859-15, UTF-8, CP1252</li><li>- ISO-8859-15, ISO-8859-1, UTF-8, CP1252</li></ul> <p>Tested under:</p> <ul style="list-style-type: none"><li>- Redmine 0.7.3.devel.2079 (MySQL)</li><li>- ruby 1.8.6 (2007-09-24 patchlevel 111) [i386-mswin32]</li><li>- Rails 2.1.0</li></ul>			
<b>Related issues:</b>			
Related to Redmine - Defect # 2664: Mercurial: Repository path encoding of no...		<b>Closed</b>	<b>2009-02-04</b>

#### Associated revisions

##### Revision 4899 - 2011-02-21 10:41 - Toshi MARUYAMA

scm: add CP932 at Setting::ENCODINGS (#2664, #2274).

CP932 is variant Japanese Shift\_JIS on Windows.

##### Revision 4906 - 2011-02-21 13:10 - Toshi MARUYAMA

scm: add "scm\_iconv" method for repository path encoding in abstract\_adapter.rb (#2664, #2274).

##### Revision 4907 - 2011-02-21 13:10 - Toshi MARUYAMA

scm: filesystem: refactor for path encoding (#2274).

**Revision 4921 - 2011-02-22 09:16 - Toshi MARUYAMA**

scm: Ruby 1.9 compatibility for browsing repository tree (#2664, #2274).

If repository path is not UTF-8, Ruby 1.9 shows trace.

**Revision 4940 - 2011-02-24 06:58 - Toshi MARUYAMA**

scm: add "path\_encoding" column in repositories table (#2664, #2274).

Contributed by Yuya Nishihara.

**Revision 4941 - 2011-02-24 06:58 - Toshi MARUYAMA**

scm: update adapter initialize() to use path encoding (#2664, #2274).

**Revision 4943 - 2011-02-24 06:59 - Toshi MARUYAMA**

scm: filesystem: add path encoding select item (#2274).

**Revision 4944 - 2011-02-24 08:00 - Toshi MARUYAMA**

scm: filesystem: fix mistake of repository select box on r4943 (#2274).

**Revision 4988 - 2011-03-02 08:11 - Toshi MARUYAMA**

scm: filesystem: add note "Default: UTF-8" in path encoding setting (#2274).

**Revision 5520 - 2011-04-23 05:35 - Toshi MARUYAMA**

scm: use i18n string at path encoding setting (#2274, #2664, #3462, #5251).

**Revision 5521 - 2011-04-23 05:35 - Toshi MARUYAMA**

scm: add Japanese translation "field\_scm\_path\_encoding" (#2274, #2664, #3462, #5251).

**Revision 5522 - 2011-04-23 05:36 - Toshi MARUYAMA**

scm: update locales for path encoding setting (#2274, #2664, #3462, #5251).

**Revision 5523 - 2011-04-23 05:36 - Toshi MARUYAMA**

scm: use i18n string at path encoding setting note (#2274, #2664, #3462, #5251).

#### **Revision 5524 - 2011-04-23 05:37 - Toshi MARUYAMA**

scm: add Japanese string at path encoding setting note (#2274, #2664, #3462, #5251).

#### **Revision 5525 - 2011-04-23 05:37 - Toshi MARUYAMA**

scm: update locales for path encoding setting note (#2274, #2664, #3462, #5251).

#### **Revision 5629 - 2011-05-04 07:02 - Toshi MARUYAMA**

scm: filesystem: fix loss non ASCII paths if path\_encoding is " (#2274).

#### **Revision 5863 - 2011-05-21 04:16 - Toshi MARUYAMA**

scm: add "path\_encoding" method in abstract adapter (#2274, #3462, #2664, #5251).

#### **Revision 5864 - 2011-05-21 04:17 - Toshi MARUYAMA**

scm: filesystem: override "path\_encoding" method in adapter (#2274).

#### **Revision 5865 - 2011-05-21 04:18 - Toshi MARUYAMA**

scm: filesystem: add unit adapter test of default path\_encoding is UTF-8 (#2274).

## **History**

---

### **#1 - 2008-12-04 20:16 - Paul Rivier**

- Assignee set to Paul Rivier

Hi Tony,

I can't reproduce it here with chars like é € à ß or à ... My only available environment is Linux, so it might be related to your windows environment. Can anybody else confirm this bug and narrow it as much as possible ?

### **#2 - 2008-12-13 20:09 - Toni Kerschbaum**

- File *repo\_enc\_bug\_01.gif* added

- File *repo\_enc\_bug\_02.gif* added

I just noticed that I forgot to upload the pictures, so here they are.

It could be related to my Windows environment (Windows Server 2003). If I can help you in any way to narrow or track down the problem, please let me know.

Also, the path in the first two pictures remains fully browseable (and the files downloadable), but some other paths and directories do not. Sadly, I can't find any notable difference in those paths and filenames.

**#3 - 2011-02-07 10:50 - Toshi MARUYAMA**

- Assignee changed from Paul Rivier to Toshi MARUYAMA

**#4 - 2011-02-08 08:11 - Toshi MARUYAMA**

- Subject changed from Filesystem Repository and (german) special chars to Filesystem Repository path encoding of non UTF-8 characters

**#5 - 2011-02-08 08:14 - Toshi MARUYAMA**

- File fs-setting.png added

- File fs-browse.png added

Try #2664 [note-19](#) patches.

These are my Japanese Windows Vista images.

**#6 - 2011-02-24 05:43 - Toshi MARUYAMA**

- Status changed from New to 7

- Target version set to 1.2.0

- % Done changed from 0 to 80

**#7 - 2011-02-24 08:04 - Toshi MARUYAMA**

- Status changed from 7 to Closed

- % Done changed from 80 to 100

- Resolution set to Fixed

I finished implementing until r4944.

It is impossible to prepare test tar ball and test non UTF-8 encoding paths on all OSs, filesystems and Languages.

If tar ball has Latin-1 path encoding files, I can't extract it on my Japanese Windows.

Please refer.

<http://mercurial.selenic.com/wiki/EncodingStrategy?action=recall&#38;rev=6>

**Files**

---

repo_enc_bug_01.gif	20.5 KB	2008-12-13	Toni Kerschbaum
repo_enc_bug_02.gif	20 KB	2008-12-13	Toni Kerschbaum
fs-setting.png	13.2 KB	2011-02-08	Toshi MARUYAMA
fs-browse.png	28.9 KB	2011-02-08	Toshi MARUYAMA