

Redmine - Feature #26279

Allow switching the encoding to UTF-8 when exporting to CSV

2017-06-27 04:09 - Go MAEDA

| | | | |
|---|----------|------------------------|-----------|
| Status: | Closed | Start date: | |
| Priority: | Normal | Due date: | |
| Assignee: | Go MAEDA | % Done: | 0% |
| Category: | Issues | Estimated time: | 0.00 hour |
| Target version: | 4.0.0 | | |
| Resolution: | Fixed | | |
| Description | | | |
| <p>In the current implementation, the encoding of exported CSV is fixed for each language and users cannot change it.</p> <p>Sometimes it is problematic for teams using multiple languages. Suppose the situation that issues in a project are written in Japanese or Chinese (mixed language project; some issue are written in Japanese and some issues are written Chinese). If you want to export those issues to CSV, using UTF-8 as the CSV encoding is only solution to get readable (not garbled) CSV file. But actually the encoding is fixed to CP932 for Japanese users (source:tags/3.3.3/config/locales/ja.yml#L164) and gb18030 for Chinese users (source:tags/3.3.3/config/locales/zh.yml#L147). No one can get CSV file in UTF-8 without modifying source code of Redmine.</p> <p>I think the problem can be resolved if users can override <code>general_csv_encoding</code> setting in CSV export options window like the following picture. The encoding in the drop-down list is defaults to <code>general_csv_encoding</code> and users can change to arbitrary encoding. We already have the similar drop-down in CSV import feature.</p> <p>csv-export-options@2x.png</p> | | | |
| Related issues: | | | |
| Related to Redmine - Defect # 27213: Euro symbol (€) is replaced by ? when pe... | | Closed | |
| Related to Redmine - Patch # 17902: CSV encoding should be UTF-8 in French lo... | | Closed | |
| Related to Redmine - Defect # 27975: CSV export of different language | | Closed | |
| Related to Redmine - Defect # 31511: CSV export of time entries report does n... | | Closed | |

Associated revisions

Revision 17328 - 2018-05-07 03:13 - Go MAEDA

Allow switching the encoding to UTF-8 when exporting to CSV (#26279).

Patch by Mizuki ISHIKAWA.

History

#1 - 2017-06-27 04:16 - Go MAEDA

- Description updated

#2 - 2017-06-27 10:05 - Felix Schäfer

Thank you for this suggestion, I think this would be a great improvement to the CSV export function.

I can corroborate the problem, we (at [Planio](#)) have had numerous customers in the last months, for example Russian users that use Planio/Redmine in English, which will have ??? in the export instead of Cyrillic characters because the export encoding for English does not support Cyrillic characters.

#3 - 2017-06-27 12:06 - Go MAEDA

Thanks, Felix. I want your advice.

At first I wrote that "users can change to arbitrary encoding". I thought that I can make a list of encodings from `Setting::ENCODINGS` constant. But now I think it is useless to display such a long list in the CSV export options and displaying only 2 choices, UTF-8 and the value of `general_csv_encoding` is enough. Because there are many incompatible encodings for each languages. For example, Japanese text is compatible only with small number of encodings such as CP932, Shift_JIS, ISO-2022-JP, EUC-JP and so on.

What do you think about this? Displaying two choices (UTF-8 and `general_csv_encoding`) are enough?

#4 - 2017-06-28 08:46 - Mizuki ISHIKAWA

I am working on this issue now. I will post a patch soon.

#5 - 2017-06-28 16:51 - Tatsuya Saito

IMHO, this option isn't needed to show EVERYTIME.

If we can change the encoding on account or system setting, almost case is probably ok.

For example of bad case, the environment is mixed several languages, and Excel on Mac which cannot recognize utf-8 CSV on opening by default.

But in this case, it will not be solved with the option on CSV exporting dialog because it will need to export with utf-8.

Any other bad case when only change the default encoding?

#6 - 2017-07-04 02:02 - Mizuki ISHIKAWA

- *File `select_encoding.patch` added*

I tried writing patch with the specification that Go MAEDA was saying.(#26279#note-3)

#7 - 2017-07-04 05:55 - Mizuki ISHIKAWA

- *File `select_encoding2.patch` added*

`select_encoding.patch` failed to apply, so I fixed it.

#8 - 2017-07-11 04:23 - Go MAEDA

- *Target version set to Candidate for next major release*

Mizuki Ishikawa, thanks for writing the patch. I tried out and it works fine on Issues and Spent time.

With the patch, we can override `general_csv_encoding` with UTF-8. "Encoding" drop-down is not displayed if `general_csv_encoding` for an user is UTF-8.

#9 - 2017-07-20 15:23 - Go MAEDA

- *Target version changed from Candidate for next major release to 4.1.0*

#10 - 2017-10-17 19:23 - Mischa The Evil

- *Related to Defect #27213: Euro symbol (€) is replaced by ? when performing a CSV export added*

#11 - 2017-10-17 19:23 - Mischa The Evil

- Related to Patch #17902: CSV encoding should be UTF-8 in French locale added

#12 - 2017-11-17 12:15 - Holger Just

This is a common issue for Redmine installations using multiple languages from different language families like English with Japanese or German with Russian. Since stored content might be mixed between the languages, it is generally very hard to find a common encoding for those. Forcing users to choose a Japanese locale in order to allow them to export Japanese text in their Redmine even if they usually use English is probably not the best user experience.

In addition to that, I understand that the current language-dependent encodings were chosen to match whatever the default encoding of Excel on Windows was for the respective language. However, this default doesn't even hold correct for other platforms. E.g. Excel 2016 for Mac with en-US locale defaults to a semicolon instead of a comma for the expected separator. In any case, both the separators as well as the encoding can be changed during opening of the file (File -> Import -> CSV file in Excel). This however requires that the exported CSV is saved with an encoding that is able to represent all of the characters in the fields.

The approach of allowing the selection of the default language and UTF-8 is in my opinion the right one. As such, I think the patch of Mizuki Ishikawa looks fine on first check since it allows to use the default encoding for the common case (as it is now) and also allows the use of UTF-8 for the more complex case. In any case, I think it's a good idea to show the user an indication of the encoding the file will be in so that they can configure their readers accordingly.

In the patch, I'd only add a check for a valid encoding in `lib/redmine/export/csv.rb` to avoid an exception in case an invalid encoding was specified.

#13 - 2018-01-12 12:58 - Go MAEDA

- Related to Defect #27975: CSV export of different language added

#14 - 2018-04-03 07:04 - Mizuki ISHIKAWA

- File `select_encoding3.patch` added

Since the patch I wrote was old, it was fixed so that it can be applied to the latest trunk (17255).

#15 - 2018-04-05 08:18 - Mizuki ISHIKAWA

- File `select_encoding4.patch` added

If unavailable encoding is selected, I have changed to use the default encoding.

#16 - 2018-05-04 04:06 - Go MAEDA

- File `select_encoding5.patch` added

Slightly updated tests.

- Made test names simpler and clearer.
- Replaced the encoding for 'encoding' option with ISO-8859-3, the encoding which is not used for `general_encoding_name` in any `config/locales/*.yml` files.

#17 - 2018-05-07 03:15 - Go MAEDA

- Subject changed from Allow to override general_csv_encoding in CSV export options window to Allow switching the encoding to UTF-8 when exporting to CSV
- Status changed from New to Closed
- Assignee set to Go MAEDA
- Target version changed from 4.1.0 to 4.0.0
- Resolution set to Fixed

Committed. Thank you for improving Redmine.

#18 - 2019-06-05 11:17 - Go MAEDA

- Related to Defect #31511: CSV export of time entries report does not honor project filter added

Files

| | | | |
|---------------------------|---------|------------|-----------------|
| csv-export-options@2x.png | 60.6 KB | 2017-06-27 | Go MAEDA |
| select_encoding.patch | 7.38 KB | 2017-07-04 | Mizuki ISHIKAWA |
| select_encoding2.patch | 7.4 KB | 2017-07-04 | Mizuki ISHIKAWA |
| select_encoding3.patch | 7.1 KB | 2018-04-03 | Mizuki ISHIKAWA |
| select_encoding4.patch | 8.48 KB | 2018-04-05 | Mizuki ISHIKAWA |
| select_encoding5.patch | 8.55 KB | 2018-05-04 | Go MAEDA |