Redmine - Feature #26423

Character encoding auto-detection using rchardet

2017-07-13 11:11 - Dmitri Zhabinski

Status:	New		**Start date:**	
Priority:	Normal		**Due date:**	
Assignee:			**% Done:**	0%
Category:	I18n		**Estimated time:**	0.00 hour
Target version:				
Resolution:				

Description

Current character encoding auto-detection cant differentiate cp1251 from cp866 (if both encodings are configured).

But https://rubygems.org/gems/rchardet/ can !

I have made proof of concept patch on my Redmine 3.3.1 from debian package.
And it just works without significant performance impact.

History

#1 - 2017-07-13 13:38 - Toshi MARUYAMA

- Category set to I18n

#2 - 2017-07-13 13:43 - Toshi MARUYAMA

Could you add tests?
source:tags/3.4.1/test/unit/lib/redmine/codeset_util_test.rb

I don't like auto detection.
Japanese has many character encodings (UTF-8, EUC-JP, CP932, Shift-JIS, and ISO-2022-JP).
Auto detection misbehaves in many cases in Japanese.

I think Redmine should have option.

- No auto detection
- unshift
- push

Files

rchardet.patch		603 Bytes	2017-07-13	Dmitri Zhabinski