

Character encoding auto-detection using rchardet

2022/05/09 18:19 - Admin Redmine

ステータス:	New	開始日:	2022/05/09
優先度:	通常	期日:	
担当者:		進捗率:	0%
カテゴリ:	l18n_37	予定工数:	0.00時間
対象バージョン:		作業時間:	0.00時間
Redmineorg_URL:	https://www.redmine.org/issues/26423	status_id:	1
category_id:	37	tracker_id:	2
version_id:	0	plus1:	0
issue_org_id:	26423	affected_version:	
author_id:	1759	closed_on:	
assigned_to_id:	0	affected_version_id:	
comments:	2		

説明

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.

journals

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

履歴

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

- カテゴリをl18n_37にセット