

Passerelle - Development #5896

créer un connecteur "nouveau style" pour datasource depuis un fichier CSV

06 novembre 2014 17:31 - Thomas Noël

Statut:	Fermé	Début:	06 novembre 2014
Priorité:	Haut	Echéance:	
Assigné à:	Serghei Mihai	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:	0.9.0	Planning:	
Patch proposed:	Oui		

Description

Reprendre le datasource CSV actuelle (éventuellement un peu from scratch, depuis un baseresource) et lui donner une nouvelle interface, dont l'édition permettrait de mettre à jour le fichier CSV.

La cible : fournir une liste de {id,text} pour wcs.

On doit fournir pour créer la ressource :

- un csv
- le rang de la colonne qui indique la clé (id) -- 0 par défaut
- le rang de la colonne qui indique le contenu (text) -- 1 par défaut

Lors de l'interrogation, un "?q=abc" renvoie uniquement les lignes qui contiennent "abc" dans le texte.

Options :

- permettre de surcharger ces rangs dans la querystring d'interrogation du webservice, genre id_row=0&text_row=3.
- text_rows=2,3,4 où le texte serait un ' '.join() des colonnes indiquées
- indiquer l'url d'un fichier CSV (qui serait mis en cache, durée du cache paramétrable...)

(note: c'est indirectement une demande cg14 pour la mise à jour du csv de dispatch APA, CSV qui sert aussi de base pour les codes postaux)

Révisions associées

Révision a017827f - 04 septembre 2015 14:24 - Serghei Mihai

csv files datasource (#5896)

Révision b7239949 - 04 septembre 2015 14:24 - Serghei Mihai

csv files datasource localizations (#5896)

Historique

#1 - 10 juin 2015 12:04 - Frédéric Péters

+ la possibilité de nommer les autres colonnes du CSV, pour utilisation en mode "étendu" de wcs.

#2 - 10 juin 2015 12:09 - Serghei Mihai

- Assigné à changé de Thomas Noël à Serghei Mihai

#3 - 30 juin 2015 19:08 - Serghei Mihai

- Fichier 0001-csv-files-data-source-5896.patch ajouté

- Statut changé de Nouveau à En cours

- Patch proposed changé de Non à Oui

Un patch brouillon.

Ça ne gère pas un csv distant et il manque les tests

#4 - 30 juin 2015 19:15 - Frédéric Péters

+ la possibilité de nommer les autres colonnes du CSV, pour utilisation en mode "étendu" de wcs.

Je voyais ça dans la définition, pas lors de l'appel.

#5 - 30 juin 2015 19:38 - Frédéric Péters

Aussi, je ne mettrais pas ce module dans contrib.

#6 - 02 juillet 2015 15:01 - Thomas Noël

Frédéric Péters a écrit :

+ la possibilité de nommer les autres colonnes du CSV, pour utilisation en mode "étendu" de wcs.

Je voyais ça dans la définition, pas lors de l'appel.

Pareil. Peut-être une simple ligne : nom,nom,nom,nom ? (pour faire vite et ne pas avoir une UI tordue)

#7 - 02 juillet 2015 16:35 - Serghei Mihai

- Fichier `csvdatasource.png` ajouté

- Fichier `0001-csv-files-data-source-5896.patch` ajouté

Perso nom, nom, nom ne me paraît pas très clair à l'usage, sauf si on y ajoute des `help_text`.
Voici l'exemple d'une UI avec un champs séparé pour les titres des colonnes

#8 - 13 juillet 2015 10:42 - Serghei Mihai

- Fichier `0001-datasource-for-csv-files-5896.patch` ajouté

Thomas Noël a écrit :

Pareil. Peut-être une simple ligne : nom,nom,nom,nom ? (pour faire vite et ne pas avoir une UI tordue)

Ok, et dans ce cas, la recherche `q=abc` ne s'appliquera que si la colonne "text" est définie.

#9 - 15 juillet 2015 17:24 - Thomas Noël

Je disais à l'instant à Serghei : pas de cache explicite du fichier CSV, comptons sur l'OS pour ça.

#10 - 15 juillet 2015 17:28 - Frédéric Péters

Oui oui, le cache c'était dans la situation d'un CSV distant (« indiquer l'url d'un fichier CSV (qui serait mis en cache, durée du cache paramétrable...) »).

#11 - 26 août 2015 11:27 - Thomas Noël

- *Priorité changé de Normal à Haut*

(ça commence à bien manquer, notamment au ministère)

#12 - 26 août 2015 15:44 - Serghei Mihai

- Fichier `0001-csv-files-datasource-5896.patch` ajouté

#13 - 26 août 2015 16:02 - Frédéric Péters

Je suis plutôt favorable à avoir un `.po` unique à Passerelle.

Pour la logique de `columns_title`. quand c'est mis, ok, on s'en sert, et quand c'est pas mis, on utilise la première ligne du CSV. J'ai l'impression que dans la pratique, ça va surtout nous ennuyer parce qu'on aura des CSV avec des titres de colonne en première ligne mais pas avec les titres qu'on veut (`id`, `text`).

```
# Create your tests here.
```

Fichier à supprimer; mais ce serait par contre super d'ajouter des tests dans un `tests/test_csv_datasource.py`.

Je me demande si on ne gagnerait pas à ne pas mélanger CSV local et CSV distant.

cache.set(self.slug, je préférerais qu'on utilise un hash de l'URL plutôt que permettre via le web d'altérer n'importe quel contenu du cache.

Le filtre, on pourra voir plus tard pour le faire évoluer.

#14 - 26 août 2015 16:37 - Serghei Mihai

Frédéric Péters a écrit :

Je suis plutôt favorable à avoir un .po unique à Passerelle.

D'accord, j'ai fait un ticket séparé: [#8121](#).

#15 - 26 août 2015 17:19 - Thomas Noël

Frédéric Péters a écrit :

Pour la logique de columns_title. quand c'est mis, ok, on s'en sert, et quand c'est pas mis, on utilise la première ligne du CSV. J'ai l'impression que dans la pratique, ça va surtout nous ennuyer parce qu'on aura des CSV avec des titres de colonne en première ligne mais pas avec les titres qu'on veut (id, text).

Il fallait faire un choix, je pense savoir guidé Serghei sur celui-ci. On peut en faire un autre, par exemple:

- une case à cocher qui indique si la première colonne est un titre, et génère des slug à partir de ça ("nom du village" -> "nom-du-village")
- dans ce cas, si columns_title existe, elle "écrase" le nom de certaines colonne : «.,id,,text» dit que "id" est dans la colonne 3 et text dans la 6ème.

Je me demande si on ne gagnerait pas à ne pas mélanger CSV local et CSV distant.

Oui, ça ferait un code beaucoup plus clair (on n'a pour l'instant pas besoin du principe des CSV distants)

Le filtre, on pourra voir plus tard pour le faire évoluer.

Oui, par rapport à <http://git.entrouvert.org/passerelle.git/tree/passerelle/datasources/models.py> on a quand même des regressions, je trouve (possibilité de choisir la/les colonne à considérer comme résultat, c'est utile pour le dispatch cg14 par exemple)

Mais on peut pousser une version simple et jouer ensuite.

#16 - 26 août 2015 17:29 - Frédéric Péters

Thomas Noël a écrit :

Frédéric Péters a écrit :

Pour la logique de columns_title. quand c'est mis, ok, on s'en sert, et quand c'est pas mis, on utilise la première ligne du CSV. J'ai l'impression que dans la pratique, ça va surtout nous ennuyer parce qu'on aura des CSV avec des titres de colonne en première ligne mais pas avec les titres qu'on veut (id, text).

Il fallait faire un choix, je pense savoir guidé Serghei sur celui-ci. On peut en faire un autre, par exemple:

- une case à cocher qui indique si la première colonne est un titre, et génère des slug à partir de ça ("nom du village" -> "nom-du-village")
- dans ce cas, si columns_title existe, elle "écrase" le nom de certaines colonne : «.,id,,text» dit que "id" est dans la colonne 3 et text dans la 6ème.

Oui, je trouve plutôt important de pouvoir zapper la première ligne.

Je me demande si on ne gagnerait pas à ne pas mélanger CSV local et CSV distant.

Oui, ça ferait un code beaucoup plus clair (on n'a pour l'instant pas besoin du principe des CSV distants)

Si on n'en a pas besoin, alors, pour moi, on peut tout à fait zapper cette partie pour l'instant et avoir du code plus simple.

#17 - 26 août 2015 17:55 - Serghei Mihai

Frédéric Péters a écrit :

Oui, je trouve plutôt important de pouvoir zapper la première ligne.

Pareil.
Et je rendrais `columns_title` obligatoire au format «`,,id,,text`», comme suggère Thomas.

#18 - 31 août 2015 12:21 - Serghei Mihai

- Fichier `0001-csv-files-datasource-5896.patch` ajouté

Avec des tests

#19 - 01 septembre 2015 12:12 - Serghei Mihai

- Version cible mis à `0.9.0`

#20 - 01 septembre 2015 16:47 - Frédéric Péters

Il n'y a plus besoin de `CACHE_SENTINEL`.

Le patch zappe désormais systématiquement la première ligne, ce n'est pas souhaité. (le test passe à côté de ça en commençant le csv par une ligne vide).

#21 - 01 septembre 2015 18:34 - Serghei Mihai

- Fichier `0001-csv-files-datasource-5896.patch` ajouté

Avec un flag pour zapper l'entete

#22 - 01 septembre 2015 22:15 - Frédéric Péters

```
titles = self.columns_titles.split(',')
indexes = [titles.index(t) for t in titles if t.split()]
caption = [titles[i] for i in indexes]
```

Il y a ce `t.split()` qui m'interloque; c'est quoi son rôle ?

```
columns_titles='id,,nom, prenom, sexe'
```

Avec une définition pareille, les noms de colonne vont garder les espaces, ça ne fera pas un truc facilement exploitable; ce serait bien que ça soit aussi testé par les tests.

Dans les tests aussi, en plus de vérifier que les dictionnaires ont les bonnes clés, ça peut être utile d'en vérifier aussi les valeurs (genre `assert csv.get_data()[0] == {'id': '69', 'text': 'DELANOUE'}`).

Dans les données de test il y a parfois des points-virgules parfois des virgules (?).

#23 - 02 septembre 2015 10:27 - Serghei Mihai

- Fichier `0001-csv-files-datasource-5896.patch` ajouté

Ça devait être un `strip`, je me suis planté.

J'ai rajouté quelques tests en plus, enlevé les virgules (qui viennent d'un export Excel en CSV)

#24 - 04 septembre 2015 10:51 - Frédéric Péters

Il manque l'appel à `gettext` :

```
class Meta:
    verbose_name = 'CSV File'
```

Et je mettrais les modifs au `.po` dans un commit séparé.

Il reste des références `local_csv_file` et `remote_csv_file` dans le template.

#25 - 04 septembre 2015 10:54 - Frédéric Péters

Aussi le `columns_titles` le "titles" m'ennuie parce qu'on y attend plutôt des identifiants (je suggère `column_keynames` à la place), et dans le `help_text` ne pas y mettre des crochets parce qu'on va se demander si on doit les taper (et je remplacerais peut-être juste par un exemple tout bête "ex: `id,text,data1,data2`").

#26 - 04 septembre 2015 11:04 - Serghei Mihai

- Fichier `0001-csv-files-datasource-5896.patch` ajouté

ça me va

#27 - 04 septembre 2015 11:09 - Serghei Mihai

- Fichier 0001-csv-files-datasource-5896.patch supprimé

#28 - 04 septembre 2015 11:28 - Serghei Mihai

- Fichier 0001-csv-files-datasource-5896.patch ajouté

- Fichier 0002-csv-files-datasource-localizations-5896.patch ajouté

Et les traductions séparées

#29 - 04 septembre 2015 12:19 - Frédéric Péters

Dernier truc j'espère, il reste un bout pas traduit; a priori ça pourrait être `{{view.model.get_verbose_name}}` :

```
<h2>CSV - {{ object.title }}</h2>
```

#30 - 04 septembre 2015 12:29 - Serghei Mihai

J'avais fait comme dans nos autres connecteurs:

```
{% block appBar %}
<h2>GDC - {{ object.title }}</h2>
...

```

ou

```
{% block appBar %}
<h2>ClicRdv - {{ object.title }}</h2>
...

```

#31 - 04 septembre 2015 12:43 - Frédéric Péters

Certes, mais d'un côté on a des marques ou noms de produit et ici on a un nom générique.

#32 - 04 septembre 2015 14:25 - Serghei Mihai

- Fichier 0001-csv-files-datasource-5896.patch ajouté

Alors pour le `get_verbose_name`

#33 - 04 septembre 2015 14:35 - Frédéric Péters

ok, go.

#34 - 04 septembre 2015 14:45 - Serghei Mihai

- Statut changé de *En cours* à *Résolu* (à déployer)

```
commit b7239949eaaf5a46ccf919846263562d8ac87767
Author: Serghei Mihai <smihai@entrouvert.com>
Date: Fri Sep 4 11:20:36 2015 +0200
```

```
csv files datasource localizations (#5896)
```

```
commit a017827f45bf90343dfdfe7d3fa0cc8678fc21f7
Author: Serghei Mihai <smihai@entrouvert.com>
Date: Wed Aug 26 15:39:01 2015 +0200
```

```
csv files datasource (#5896)
```

#35 - 27 mars 2017 11:08 - Serghei Mihai

- Statut changé de *Résolu* (à déployer) à *Fermé*

Fichiers

0001-csv-files-data-source-5896.patch

15,3 ko

30 juin 2015

Serghei Mihai

csvdatasource.png	34,6 ko	02 juillet 2015	Serghei Mihai
0001-csv-files-data-source-5896.patch	16,2 ko	02 juillet 2015	Serghei Mihai
0001-datasource-for-csv-files-5896.patch	16,2 ko	13 juillet 2015	Serghei Mihai
0001-csv-files-datasource-5896.patch	19 ko	26 août 2015	Serghei Mihai
0001-csv-files-datasource-5896.patch	20,1 ko	31 août 2015	Serghei Mihai
0001-csv-files-datasource-5896.patch	20,6 ko	01 septembre 2015	Serghei Mihai
0001-csv-files-datasource-5896.patch	20,9 ko	02 septembre 2015	Serghei Mihai
0001-csv-files-datasource-5896.patch	15,4 ko	04 septembre 2015	Serghei Mihai
0002-csv-files-datasource-localizations-5896.patch	5,29 ko	04 septembre 2015	Serghei Mihai
0001-csv-files-datasource-5896.patch	15,7 ko	04 septembre 2015	Serghei Mihai