

Authentic 2 - Bug #37374

import csv, afficher un message en cas de mauvais type de fichier

31 octobre 2019 09:17 - Frédéric Péters

Statut:	Fermé	Début:	31 octobre 2019
Priorité:	Normal	Echéance:	
Assigné à:	Valentin Deniaud	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	Non
Patch proposed:	Oui		
Description			
Aujourd'hui,			
File <code>"/usr/lib/python2.7/dist-packages/authentic2/csv_import.py"</code> in run			
137. <code>return parse_csv()</code>			
File <code>"/usr/lib/python2.7/dist-packages/authentic2/csv_import.py"</code> in parse_csv			
119. <code>dialect = csv.Sniffer().sniff(input_fd.read().encode('utf-8'))</code>			
File <code>"/usr/lib/python2.7/codecs.py"</code> in decode			
314. <code>(result, consumed) = self._buffer_decode(data, self.errors, final)</code>			
Exception Type: UnicodeDecodeError at /manage/users/import/			
Exception Value: 'utf8' codec can't decode byte 0xe9 in position 20: invalid continuation byte [...]			
FILES:			
import_file = <InMemoryUploadedFile: 59RSA93_1909010_0.csv (application/vnd.ms-excel)>			

Révisions associées

Révision 2936f256 - 21 novembre 2019 10:20 - Valentin Deniaud

csv_import: display error message on bad encoding (#37374)

Révision fba5f121 - 21 novembre 2019 10:20 - Valentin Deniaud

manager: clearer error message on invalid csv import file (#37374)

Historique

#1 - 19 novembre 2019 14:59 - Valentin Deniaud

- Fichier `0001-csv_import-display-error-message-on-bad-encoding-373.patch` ajouté
- Statut changé de Nouveau à Solution proposée
- Patch proposed changé de Non à Oui

C'est un bug d'encodage à la lecture (le fait que ça soit un csv valide ou pas est traité ensuite). Donc je le corrige comme tel, et je n'implémente pas de vérification du mimetype comme pouvait le laisser sous entendre le titre du ticket.

#2 - 19 novembre 2019 14:59 - Valentin Deniaud

- Assigné à mis à Valentin Deniaud

#3 - 19 novembre 2019 16:19 - Frédéric Péters

je n'implémente pas de vérification du mimetype comme pouvait le laisser sous entendre le titre du ticket.

L'idée sous-entendue était que si le type MIME était whatever/excel, il y ait un message pour dire "chouchou tape un CSV".

#4 - 19 novembre 2019 16:36 - Valentin Deniaud

« Mais mon fichier s'appelle 59RSA93_1909010_0.csv, qu'est-ce que c'est que cette erreur qui me dit que c'est pas un CSV ? » --> erreur d'encodage ou erreur de type, c'est kif kif niveau compréhension je trouve.
Et puis comme on ne va pas refuser un fichier dont le mime n'est pas text/csv, il faudrait pour de vrai établir une liste de types à exclure, ça sera jamais propre imo. Je préfère m'en tenir à fixer la trace.

En passant j'aimerais bien savoir comment Django se retrouve informé du vrai mimetype, sûrement par la requête http, mais du coup quel navigateur s'amuse à parser le header d'un fichier avant de l'uploader ? Pas firefox j'ai l'impression.

#5 - 19 novembre 2019 17:45 - Benjamin Dauvergne

Valentin Deniaud a écrit :

« Mais mon fichier s'appelle 59RSA93_1909010_0.csv, qu'est-ce que c'est que cette erreur qui me dit que c'est pas un CSV ? » --> erreur d'encodage ou erreur de type, c'est kif kif niveau compréhension je trouve.
Et puis comme on ne va pas refuser un fichier dont le mime n'est pas text/csv, il faudrait pour de vrai établir une liste de types à exclure, ça sera jamais propre imo. Je préfère m'en tenir à fixer la trace.

Oui c'est plus simple je pense, le mime-type envoyé n'est jamais fiable.

En passant j'aimerais bien savoir comment Django se retrouve informé du vrai mimetype, sûrement par la requête http, mais du coup quel navigateur s'amuse à parser le header d'un fichier avant de l'uploader ? Pas firefox j'ai l'impression.

Sur Firefox c'est principalement basé sur l'extension¹.

¹https://developer.mozilla.org/en-US/docs/Mozilla/How_Mozilla_determines_MIME_Types#ExternalHelperAppService

#6 - 19 novembre 2019 17:52 - Benjamin Dauvergne

Après ça pourrait être sympa de faire la différence entre un CSV en latin1, latin15 ou CP-1252 et un fichier binaire; je n'ai pas réellement d'idée sur comment faire sachant que pour les encodages 8bits (latin1&co) .decode('latin1') marchera toujours.

On pourrait en profiter aussi pour rationaliser un poil les messages d'erreur genre "Fichier d'import invalide: <un truc plus précis>".

PS: une piste sur stackoverflow <https://stackoverflow.com/questions/1446549/how-to-identify-binary-and-text-files-using-python>

#7 - 20 novembre 2019 17:13 - Valentin Deniaud

L'histoire des encodages c'est beaucoup de hack pour pas grand chose, je suis pour ouvrir un autre ticket dev si nécessaire et laisser celui-ci être une simple correction de bug (et je suggérerais ensuite d'utiliser filemagic et ainsi de se laver les mains de tous les faux positifs qu'on risque de se prendre).

On pourrait en profiter aussi pour rationaliser un poil les messages d'erreur genre "Fichier d'import invalide: <un truc plus précis>".

Je fais ça dans import_csv.py ou au moment de l'affichage dans forms.py ?

#8 - 20 novembre 2019 19:19 - Benjamin Dauvergne

Valentin Deniaud a écrit :

L'histoire des encodages c'est beaucoup de hack pour pas grand chose, je suis pour ouvrir un autre ticket dev si nécessaire et laisser celui-ci être une simple correction de bug (et je suggérerais ensuite d'utiliser filemagic et ainsi de se laver les mains de tous les faux positifs qu'on risque de se prendre).

Ok.

On pourrait en profiter aussi pour rationaliser un poil les messages d'erreur genre "Fichier d'import invalide: <un truc plus précis>".

Je fais ça dans import_csv.py ou au moment de l'affichage dans forms.py ?

Das forms.py je pense.

#9 - 21 novembre 2019 10:23 - Valentin Deniaud

- Fichier 0002-manager-clearer-error-message-on-invalid-csv-import-patch ajouté
- Fichier 0001-csv_import-display-error-message-on-bad-encoding-373.patch ajouté

Voilà, manquera juste les traductions.

#10 - 21 novembre 2019 10:59 - Benjamin Dauvergne

- Statut changé de Solution proposée à Solution validée

#11 - 21 novembre 2019 15:09 - Valentin Deniaud

- Statut changé de Solution validée à Résolu (à déployer)

```
commit fba5f1217cc82d2373a8e84b6d79f7c2d860ca95
Author: Valentin Deniaud <vdeniaud@entrouvert.com>
Date: Thu Nov 21 10:20:34 2019 +0100
```

```
manager: clearer error message on invalid csv import file (#37374)
```

```
commit 2936f25623e137f9a805cabdf0f4aa002e534118
Author: Valentin Deniaud <vdeniaud@entrouvert.com>
Date: Tue Nov 19 14:42:17 2019 +0100
```

```
csv_import: display error message on bad encoding (#37374)
```

#12 - 04 décembre 2019 18:15 - Frédéric Péters

- Statut changé de Résolu (à déployer) à Solution déployée

Fichiers

0001-csv_import-display-error-message-on-bad-encoding-373.patch	1,97 ko	19 novembre 2019	Valentin Deniaud
0002-manager-clearer-error-message-on-invalid-csv-import-.patch	2,19 ko	21 novembre 2019	Valentin Deniaud
0001-csv_import-display-error-message-on-bad-encoding-373.patch	1,97 ko	21 novembre 2019	Valentin Deniaud