

Passerelle - Bug #13749

connecteur CSV : meilleur comportement sur des données non-UTF-8

26 octobre 2016 09:48 - Frédéric Péters

Statut:	Nouveau	Début:	26 octobre 2016
Priorité:	Normal	Echéance:	
Assigné à:		% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	
Patch proposed:	Non		
Description			
Aujourd'hui c'est erreur 500.			
Internal Server Error: /csvdatasource/liste-des-rues/data			
Traceback (most recent call last):			
File "/usr/lib/python2.7/dist-packages/django/core/handlers/base.py", line 111, in get_response			
response = wrapped_callback(request, *callback_args, **callback_kwargs)			
File "/usr/lib/python2.7/dist-packages/passerelle/urls_utils.py", line 54, in f			
return func(request, *args, **kwargs)			
File "/usr/lib/python2.7/dist-packages/django/views/generic/base.py", line 69, in view			
return self.dispatch(request, *args, **kwargs)			
File "/usr/lib/python2.7/dist-packages/django/views/generic/base.py", line 87, in dispatch			
return handler(request, *args, **kwargs)			
File "/usr/lib/python2.7/dist-packages/passerelle/urls/__init__.py", line 104, in _wrapped_view			
return view_func(instance, request, *args, **kwargs)			
File "/usr/lib/python2.7/dist-packages/passerelle/urls/jsonresponse.py", line 293, in wrapper			
return self.method(f, *args, **kwargs)			
File "/usr/lib/python2.7/dist-packages/passerelle/urls/jsonresponse.py", line 359, in api_method			
return self.api(f, args[1], *args, **kwargs)			
File "/usr/lib/python2.7/dist-packages/passerelle/urls/jsonresponse.py", line 401, in api			
return self.render_data(req, data, status, data['err'])			
File "/usr/lib/python2.7/dist-packages/passerelle/urls/jsonresponse.py", line 345, in render_data			
plain = json.dumps(data, **kwargs)			
File "/usr/lib/python2.7/json/__init__.py", line 238, in dumps			
**kw).encode(obj)			
File "/usr/lib/python2.7/json/encoder.py", line 200, in encode			
chunks = self.iterencode(o, _one_shot=True)			
File "/usr/lib/python2.7/json/encoder.py", line 263, in iterencode			
return _iterencode(o, 0)			
UnicodeDecodeError: 'utf8' codec can't decode byte 0xe9 in position 22: unexpected end of data			

Historique

#1 - 26 octobre 2016 10:57 - Benjamin Dauvergne

Je dirai que le mieux ce serait de les refuser au dépôt et pour couvrir le passé de faire une validation aussi avant le traitement de chaque requête (faut traiter /data et /query, mais /query utiliser un UnicodeDictReader qu'on pourrait améliorer pour faire cette détection et aussi utiliser dans /data).

#2 - 26 octobre 2016 10:59 - Benjamin Dauvergne

On pourrait aussi utiliser 'replace' au lieu 'ignore' dans

```
content = content.decode('utf-8-sig', 'ignore').encode('utf-8')
```

bon ça cache les erreurs c'est vrai.