

django-mellon - Development #10196

Quand un fichier de métadonnée est issue d'une URL ou d'un fichier il n'est jamais plus rechargé

04 mars 2016 10:26 - Benjamin Dauvergne

Statut:	Fermé	Début:	04 mars 2016
Priorité:	Normal	Echéance:	
Assigné à:	Benjamin Dauvergne	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	Non
Patch proposed:	Oui		

Description

Il faudrait conserver un timestamp de la dernière lecture/récupération des métadonnées, et dans le cas d'un fichier conserver le chemin (peut-être en changeant le nom de la clé en METADATA_FILE).

On devrait récupérer le fichier et mettre à jour les métadonnées quand:

- le timestamp du fichier est plus récent que celui sauvegardé
- les métadonnées précise cacheDuration et le dernier timestamp est plus vieux que cela
- les métadonnées précise validUntil qui est plus récent que le dernier timestamp mais inférieur au moment présent (si le validUntil était déjà obsolète lors du dernier chargement, on l'ignore)

Si la mise à jour ou la récupération échoue, on continue avec les métadonnées en cache.

Révisions associées

Révision 74b61de6 - 11 avril 2016 19:14 - Benjamin Dauvergne

replace dateutil by isodate (#10196)

isodate has better support for the full ISO8601 specification.

Révision 968aa07f - 18 juin 2019 00:53 - Benjamin Dauvergne

really retrieve XML encoding (#10196)

Révision 83a09d87 - 18 juin 2019 00:53 - Benjamin Dauvergne

code style (#10196)

Révision 39e2e7e5 - 18 juin 2019 00:53 - Benjamin Dauvergne

update and cache metadata from URL and path (#10196)

Historique

#1 - 11 avril 2016 20:44 - Frédéric Péters

Ticket sans rapport ?

#2 - 11 avril 2016 22:47 - Benjamin Dauvergne

Si je vais avoir besoin d'isodate pour parser l'attribut cacheDuration et de toute façon isodate m'inspire plus confiance que dateutil pour parser des dates iso.

#3 - 27 avril 2016 10:01 - Benjamin Dauvergne

- Version cible 1.2.25 supprimé

#4 - 27 avril 2016 10:01 - Benjamin Dauvergne

- Assigné à mis à Benjamin Dauvergne

#5 - 27 septembre 2017 14:31 - Benjamin Dauvergne

- Assigné à Benjamin Dauvergne supprimé

#6 - 07 juin 2019 14:50 - Benjamin Dauvergne

- Assigné à mis à Benjamin Dauvergne

#7 - 07 juin 2019 14:50 - Benjamin Dauvergne

- Fichier 0002-update-and-cache-metadata-from-URL-and-path-10196.patch ajouté

- Fichier 0001-code-style-10196.patch ajouté

- Statut changé de Nouveau à Solution proposée

- Patch proposed changé de Non à Oui

atomicwrites est dans stretch j'ai vérifié.

#8 - 07 juin 2019 15:16 - Frédéric Péters

Mais ça veut dire ne plus builder pour jessie alors. (moi ça me va)

#9 - 07 juin 2019 15:17 - Frédéric Péters

(ou taper ce paquet dans notre dépôt jessie, sans doute jouable)

#10 - 07 juin 2019 15:26 - Frédéric Péters

(ou taper ce paquet dans notre dépôt jessie, sans doute jouable)

fait.

#11 - 07 juin 2019 16:50 - Benjamin Dauvergne

- Fichier 0002-update-and-cache-metadata-from-URL-and-path-10196.patch ajouté

- Fichier 0001-code-style-10196.patch ajouté

Avec un peu plus de contenu pour le README notamment pour MELLON_IDENTITY_PROVIDERS.

Pour rétrocompatibilité je supporte toujours METADATA=chemin mais il faudrait migrer (notamment le setting loader dans hobo) vers la clé METADATA_PATH.

#12 - 07 juin 2019 18:07 - Benjamin Dauvergne

- Fichier 0002-update-and-cache-metadata-from-URL-and-path-10196.patch ajouté

- Fichier 0001-code-style-10196.patch ajouté

et une alerte si des métadonnées ne sont pas mises à jour pendant plus de 24 * CACHE_TIME (donc 1 jour par défaut).

#13 - 07 juin 2019 21:49 - Benjamin Dauvergne

- Fichier 0002-update-and-cache-metadata-from-URL-and-path-10196.patch ajouté

- Fichier 0001-code-style-10196.patch ajouté

Un dernier coup pour éviter encore plus de logs inutiles.

#14 - 17 juin 2019 12:58 - Christophe Siraut

Je reformulerais: All other keys are override of generic settings -> All other keys override generic settings.

La logique et l'implémentation me semblent ok. Est-ce qu'on anticipe qu'il y'aura éventuellement des installations qui ne fonctionneront plus parce que URL distante absente et durée de cache dépassée?

#15 - 17 juin 2019 14:21 - Frédéric Péters

URL distante absente et durée de cache dépassée

Le cache était dans la mémoire des processus, qui s'éteignaient de toute façon avec le temps etc.

#16 - 17 juin 2019 14:32 - Thomas Noël

Concernant 0001-code-style-10196.patch, sur mellon/utils.py je suis curieux, pourquoi passer par un xml_encoding[0] au lieu du simple xml_encoding précédent ?

Dans le même patch, concernant le expect après le isodate.parse_datetime, on couvre actuellement toutes les exceptions, tu as restreint ici à ValueError, TypeError mais on pourrait en avoir d'autres si date_string n'est pas une string par exemple. Je serais pour être bête et laisser un "except Exception:" général. (Parce qu'on traite ici des bouts d'objets réponses de lasso, et je suis pas à l'aise pour savoir si c'est toujours des strings)

(0002 non relu encore par moi, je laisse Christophe finir ou je peux venir en rerelecteur si besoin)

#17 - 17 juin 2019 18:16 - Benjamin Dauvergne

Thomas Noël a écrit :

Concernant 0001-code-style-10196.patch, sur mellon/utils.py je suis curieux, pourquoi passer par un xml_encoding[0] au lieu du simple xml_encoding précédent ?

Dans le même patch, concernant le expect après le isodate.parse_datetime, on couvre actuellement toutes les exceptions, tu as restreint ici à ValueError, TypeError mais on pourrait en avoir d'autres si date_string n'est pas une string par exemple. Je serais pour être bête et laisser un "except Exception:" général. (Parce qu'on traite ici des bouts d'objets réponses de lasso, et je suis pas à l'aise pour savoir si c'est toujours des strings)

Pour moi si c'est pas une string ça fera forcément TypeError (derrière ça utilise re je pense qui balance du TypeError dès qu'on a pas une string); et en vrai lasso renvoie juste une string ou None ou vraiment au pire un LassoNode (un objet Python tout con). Même pas peur.

(0002 non relu encore par moi, je laisse Christophe finir ou je peux venir en rerelecteur si besoin)

#18 - 17 juin 2019 18:17 - Benjamin Dauvergne

- Fichier 0001-really-retrieve-XML-encoding-10196.patch ajouté

- Fichier 0002-code-style-10196.patch ajouté

- Fichier 0003-update-and-cache-metadata-from-URL-and-path-10196.patch ajouté

rebasé sur master, j'ai séparé les histoires de xml_encoding, en vrai même si c'est None ça marche (parce que si on met response.encoding à None, requests autodétecte ou prend UTF-8 par défaut je ne sais pas trop, en tout cas les tests passent quand même avec un <?xml?> en plus dans les réponses SAML).

#19 - 18 juin 2019 00:38 - Thomas Noël

Benjamin Dauvergne a écrit :

Thomas Noël a écrit :

Concernant 0001-code-style-10196.patch, sur mellon/utils.py je suis curieux, pourquoi passer par un xml_encoding[0] au lieu du simple xml_encoding précédent ?

Pourquoi t'as pas répondu à ma question métaphysique d'abord hein ?

Dans le même patch, concernant le expect après le isodate.parse_datetime, on couvre actuellement toutes les exceptions, tu as restreint ici à ValueError, TypeError mais on pourrait en avoir d'autres si date_string n'est pas une string par exemple. Je serais pour être bête et laisser un "except Exception:" général. (Parce qu'on traite ici des bouts d'objets réponses de lasso, et je suis pas à l'aise pour savoir si c'est toujours des strings)

Pour moi si c'est pas une string ça fera forcément TypeError (derrière ça utilise re je pense qui balance du TypeError dès qu'on a pas une string); et en vrai lasso renvoie juste une string ou None ou vraiment au pire un LassoNode (un objet Python tout con). Même pas peur.

Je disais ça parce que justement, None :

```
isodate.parse_datetime(None) -> AttributeError: 'NoneType' object has no attribute 'split'
```

#20 - 18 juin 2019 00:49 - Benjamin Dauvergne

Thomas Noël a écrit :

Benjamin Dauvergne a écrit :

Thomas Noël a écrit :

Concernant 0001-code-style-10196.patch, sur mellon/utis.py je suis curieux, pourquoi passer par un xml_encoding[0] au lieu du simple xml_encoding précédent ?

Pourquoi t'as pas répondu à ma question métaphysique d'abord hein ?

Ah.. ma réponse s'est perdue; en python tu ne peux pas modifier dans une closure une variable du scope supérieur, elle devient automatiquement locale; le simple fait d'avoir une affectation dans un scope rend la variable locale. Mais j'aurai pu faire un global xml_encoding aussi, je ne sais pas pourquoi je n'aime pas beaucoup.

Je disais ça parce que justement, None :

```
isodate.parse_datetime(None) -> AttributeError: 'NoneType' object has no attribute 'split'
```

Ok alors.

#21 - 18 juin 2019 00:54 - Benjamin Dauvergne

- Fichier 0001-really-retrieve-XML-encoding-10196.patch ajouté
- Fichier 0002-code-style-10196.patch ajouté
- Fichier 0003-update-and-cache-metadata-from-URL-and-path-10196.patch ajouté

Poum.

#22 - 18 juin 2019 09:05 - Christophe Siraut

- Statut changé de Solution proposée à Solution validée

#23 - 18 juin 2019 15:29 - Benjamin Dauvergne

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

```
commit 39e2e7e5acaf390bbb93f812560d870799f18c1e
Author: Benjamin Dauvergne <bdauvergne@entrouvert.com>
Date: Fri Jun 7 21:46:07 2019 +0200
```

```
update and cache metadata from URL and path (#10196)
```

```
commit 83a09d874e6afda1b128d2b8b5eacb5b026244d1
Author: Benjamin Dauvergne <bdauvergne@entrouvert.com>
Date: Fri Jun 7 21:43:23 2019 +0200
```

```
code style (#10196)
```

```
commit 968aa07faf7110e32d36252f30d7cac47f8c215e
Author: Benjamin Dauvergne <bdauvergne@entrouvert.com>
Date: Mon Jun 17 18:13:09 2019 +0200
```

```
really retrieve XML encoding (#10196)
```

#24 - 18 juin 2019 16:16 - Frédéric Péters

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

Fichiers

0002-update-and-cache-metadata-from-URL-and-path-10196.patch	18,1 ko	07 juin 2019	Benjamin Dauvergne
0001-code-style-10196.patch	10,9 ko	07 juin 2019	Benjamin Dauvergne
0002-update-and-cache-metadata-from-URL-and-path-10196.patch	19,6 ko	07 juin 2019	Benjamin Dauvergne
0001-code-style-10196.patch	10,9 ko	07 juin 2019	Benjamin Dauvergne
0002-update-and-cache-metadata-from-URL-and-path-10196.patch	19,6 ko	07 juin 2019	Benjamin Dauvergne
0001-code-style-10196.patch	10,9 ko	07 juin 2019	Benjamin Dauvergne
0002-update-and-cache-metadata-from-URL-and-path-10196.patch	33,5 ko	07 juin 2019	Benjamin Dauvergne
0001-code-style-10196.patch	18,8 ko	07 juin 2019	Benjamin Dauvergne
0001-really-retrieve-XML-encoding-10196.patch	1,4 ko	17 juin 2019	Benjamin Dauvergne
0002-code-style-10196.patch	18,4 ko	17 juin 2019	Benjamin Dauvergne

0003-update-and-cache-metadata-from-URL-and-path-10196.patch	38,5 ko	17 juin 2019	Benjamin Dauvergne
0001-really-retrieve-XML-encoding-10196.patch	1,47 ko	17 juin 2019	Benjamin Dauvergne
0002-code-style-10196.patch	18,4 ko	17 juin 2019	Benjamin Dauvergne
0003-update-and-cache-metadata-from-URL-and-path-10196.patch	38,1 ko	17 juin 2019	Benjamin Dauvergne