Zoo - Development #37038

nanterre: pénaliser les frères/soeurs et les conjoints lors de la recherche de doublon

17 octobre 2019 14:52 - Benjamin Dauvergne

Statut:	Fermé	**Début:**	17 octobre 2019
Priorité:	Normal	**Echéance:**	
Assigné à:	Benjamin Dauvergne	**% réalisé:**	0%
Catégorie:		**Temps estimé:**	0:00 heure
Version cible:			
Patch proposed:	Oui	**Planning:**	Non

Description

Ils sont forcément proches (nom et date de naissance identique pour les jumeaux, nom pour les fratries et les conjoints).

Révisions associées

Révision fd0ca4b6 - 15 janvier 2020 23:48 - Benjamin Dauvergne

nanterre: prévient les erreurs unicode quand on pipe rsu-duplicates (#37038)

Révision f7183bd2 - 15 janvier 2020 23:50 - Benjamin Dauvergne

nanterre: diminuer la similarité pour les membres d'une même famille (#37038)

Historique

#2 - 09 janvier 2020 17:55 - Benjamin Dauvergne

- Tracker changé de Support à Development

#3 - 09 janvier 2020 17:56 - Benjamin Dauvergne

- Assigné à mis à Benjamin Dauvergne

#4 - 12 janvier 2020 21:58 - Benjamin Dauvergne

- Fichier 0001-nanterre-pr-vient-les-erreurs-unicode-quand-on-pipe-.patch ajouté

- Fichier 0002-nanterre-diminuer-la-similarit-pour-les-membres-d-un.patch ajouté

- Statut changé de Nouveau à Solution proposée

- Patch proposed changé de Non à Oui

#5 - 13 janvier 2020 08:27 - Emmanuel Cazenave

Tes messages de commit sont originaux.

#6 - 13 janvier 2020 10:55 - Benjamin Dauvergne

Emmanuel Cazenave a écrit :

> Tes messages de commit sont originaux.

Ouais j'ai pris l'habitude de les écrire en français parce que ça ne concerne que Nanterre.

#7 - 13 janvier 2020 11:16 - Emmanuel Cazenave

Benjamin Dauvergne a écrit :

> Ouais j'ai pris l'habitude de les écrire en français parce que ça ne concerne que Nanterre.

Ah c'est les messages dans les patchs joints qui sont foireux, dans ta branche c'est ok, bizarre.

Pas compris ça, où ce queryset est-il utilisé ?

```
# clear old duplicates
Duplicate.objects.filter(id__in=[known[p].id for p in set(known) - set(seen)])
```

#8 - 13 janvier 2020 11:31 - Benjamin Dauvergne

Emmanuel Cazenave a écrit :

> Pas compris ça, où ce queryset est-il utilisé ?
>
> [...]

Bah nulle part, j'ai oublié le .delete() merci :) Ça souligne juste le manque de tests sur cette fonctionnalité.

#9 - 13 janvier 2020 11:31 - Benjamin Dauvergne

- Statut changé de Solution proposée à En cours

#10 - 13 janvier 2020 11:49 - Benjamin Dauvergne

- Fichier 0001-nanterre-pr-vient-les-erreurs-unicode-quand-on-pipe-.patch ajouté

- Fichier 0002-nanterre-diminuer-la-similarit-pour-les-membres-d-un.patch ajouté

- Statut changé de En cours à Solution proposée

Avec un test c'est plus sympa.

#11 - 13 janvier 2020 13:56 - Benjamin Dauvergne

- Fichier 0001-nanterre-pr-vient-les-erreurs-unicode-quand-on-pipe-.patch ajouté

- Fichier 0002-nanterre-diminuer-la-similarit-pour-les-membres-d-un.patch ajouté

#12 - 13 janvier 2020 16:13 - Thomas Noël

Sur 0001 tu peux en profiter pour virer le "import sys"

Dans 0002 pourquoi tu utilises "limit or ..." pour sibling_factor ? Je trouve qu'un simple « sibling_factor = getattr(settings, 'ZOO_NANTERRE_DUPLICATES_SIBLING_FACTOR', 0.6) » serait plus explicite.

Détail, ligne 56 y'a une espace après conjoints dans « conjoints .add(...) »

#13 - 15 janvier 2020 23:51 - Benjamin Dauvergne

- Fichier 0001-nanterre-pr-vient-les-erreurs-unicode-quand-on-pipe-.patch ajouté

- Fichier 0002-nanterre-diminuer-la-similarit-pour-les-membres-d-un.patch ajouté

> Dans 0002 pourquoi tu utilises "limit or ..." pour sibling_factor ? Je trouve qu'un simple « sibling_factor = getattr(settings, 'ZOO_NANTERRE_DUPLICATES_SIBLING_FACTOR', 0.6) » serait plus explicite.

Un copier/coller malheureux, j'ai corrigé.

#14 - 16 janvier 2020 11:16 - Thomas Noël

- Statut changé de Solution proposée à Solution validée

#15 - 16 janvier 2020 15:46 - Benjamin Dauvergne

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

```
commit f7183bd2ce991ef52e71d273819cc71e9a98063e
Author: Benjamin Dauvergne <bdauvergne@entrouvert.com>
Date:   Sun Jan 12 21:57:22 2020 +0100

    nanterre: diminuer la similarité pour les membres d'une même famille (#37038)

commit fd0ca4b6aa35e499d9bf97df3e9b7b9638ebd1ad
Author: Benjamin Dauvergne <bdauvergne@entrouvert.com>
Date:   Fri Jan 10 08:19:33 2020 +0100

    nanterre: prévient les erreurs unicode quand on pipe rsu-duplicates (#37038)
```

#16 - 05 janvier 2022 20:39 - Frédéric Péters

- Statut changé de Résolu (à déployer) à Fermé

Fichiers

0001-nanterre-pr-vient-les-erreurs-unicode-quand-on-pipe-.patch	4 ko	12 janvier 2020	Benjamin Dauvergne
0002-nanterre-diminuer-la-similarit-pour-les-membres-d-un.patch	3,89 ko	12 janvier 2020	Benjamin Dauvergne
0001-nanterre-pr-vient-les-erreurs-unicode-quand-on-pipe-.patch	4 ko	13 janvier 2020	Benjamin Dauvergne
0002-nanterre-diminuer-la-similarit-pour-les-membres-d-un.patch	5,21 ko	13 janvier 2020	Benjamin Dauvergne
0001-nanterre-pr-vient-les-erreurs-unicode-quand-on-pipe-.patch	4 ko	13 janvier 2020	Benjamin Dauvergne
0002-nanterre-diminuer-la-similarit-pour-les-membres-d-un.patch	5,3 ko	13 janvier 2020	Benjamin Dauvergne
0001-nanterre-pr-vient-les-erreurs-unicode-quand-on-pipe-.patch	4 ko	15 janvier 2020	Benjamin Dauvergne
0002-nanterre-diminuer-la-similarit-pour-les-membres-d-un.patch	5,3 ko	15 janvier 2020	Benjamin Dauvergne