

Zoo - Development #19092

nanterre: améliorer l'indexation des noms complets des individus

29 septembre 2017 12:30 - Benjamin Dauvergne

Statut:	Fermé	Début:	29 septembre 2017
Priorité:	Normal	Echéance:	
Assigné à:	Benjamin Dauvergne	% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	
Patch proposed:	Non		
Description			
Actuellement on indexe individuellement nom et prénom et on essaie de recalculer une valeur de similarité à partir des similarités partielles ce qui n'est pas idéal.			
L'amélioration serait de créer deux index trigram un sur:			
<pre>NORMALIZE (CONCAT (prenoms, ' ', nom_de_naissance))</pre>			
le deuxième sur			
<pre>NORMALIZE (CONCAT (prenoms, ' ', nom_d_usage))</pre>			
On arrêterait d'essayer de chercher le nom ou le prénom séparément on chercherait directement la chaîne complète prénoms/nom dans les deux index ce qui augmenterait la pertinence des recherches.			
Le travail est en cours sur la branche wip/improve-name-search : http://git.entrouvert.org/zoo.git/log/?h=wip/improve-name-search			
Demandes liées:			
Lié à Zoo - Bug #21619: zoo nanterre: ne démarre plus parce que le migrate la...		Fermé	02 février 2018
Lié à Zoo - Bug #21620: crash sur détection de doublons		Fermé	02 février 2018
Bloque Zoo - Development #21415: nanterre: rendre plus souple la recherche pa...		Fermé	25 janvier 2018

Révisions associées

Révision 08846955 - 02 février 2018 18:03 - Benjamin Dauvergne

zoo_meta: reconstruit les index via les signaux (#19092)

Soit après un signal post_save sur un schéma soit après une migration.

Révision efaafa57 - 02 février 2018 18:03 - Benjamin Dauvergne

zoo_meta: ajoute une méthode utilitaire pour créer des index trigram (#19092)

Révision 6137d25a - 02 février 2018 18:03 - Benjamin Dauvergne

zoo_meta: ajoute un hook sur les classes AppConfig pour personnaliser la reconstruction des index (#19092)

Révision 86841398 - 02 février 2018 18:03 - Benjamin Dauvergne

zoo_data: ajoute une nouvelle fonction SQL TextCat (#19092)

Pour remplacer les utilisations de CONCAT dans les conditions WHERE qui ne peut donc pas être indexé, contrairement à l'opérateur || classique.

Révision 31f1d2b3 - 02 février 2018 18:03 - Benjamin Dauvergne

nanterre: construire des index spécifiques pour les noms complets des individus (#19092)

Deux index sont construits:

- prenoms + noms_de_naissance
- prenoms + nom_d_usage

Révision 8f02fb31 - 02 février 2018 18:03 - Benjamin Dauvergne

nanterre: améliore la recherche par nom (fixes #19092)

En utilisant les index sur le nom complet plutôt que sur le prénom ou le nom.

Historique

#4 - 01 février 2018 15:55 - Benjamin Dauvergne

- Fichier 0005-nanterre-construire-des-index-sp-cifiques-pour-les-n.patch ajouté
- Fichier 0002-zoo_meta-ajoute-une-m-thode-utilitaire-pour-cr-er-de.patch ajouté
- Fichier 0003-zoo_meta-ajoute-un-hook-sur-les-classes-AppConfig-po.patch ajouté
- Fichier 0001-zoo_meta-reconstruit-les-index-via-les-signaux-19092.patch ajouté
- Fichier 0004-zoo_data-ajoute-une-nouvelle-fonction-SQL-TextCat-19.patch ajouté
- Fichier 0006-nanterre-am-liore-la-recherche-par-nom-fixes-19092.patch ajouté

Les 4 premiers patches améliorent l'infrastructure zoo en vue des deux derniers qui crée de meilleurs index et s'en servent dans la recherche.

J'en profite pour réorganiser la classe de recherche, notamment en rendant plus générique l'aggrégation des filtres et des expressions de score.

#5 - 02 février 2018 16:28 - Benjamin Dauvergne

- Lié à Development #21415: nanterre: rendre plus souple la recherche par date de naissance ajouté

#6 - 02 février 2018 16:28 - Benjamin Dauvergne

- Lié à Development #21415: nanterre: rendre plus souple la recherche par date de naissance supprimé

#7 - 02 février 2018 16:28 - Benjamin Dauvergne

- Bloque Development #21415: nanterre: rendre plus souple la recherche par date de naissance ajouté

#8 - 02 février 2018 18:02 - Thomas Noël

0001 à 0005: ack

0006 : on voit que les patches précédents permettent un nettoyage du code ici, j'aime bien. Rien vu d'étrange ou incompréhensible.

Ack et j'envoie en recette ASAP.

#9 - 02 février 2018 18:05 - Benjamin Dauvergne

- Statut changé de Nouveau à Résolu (à déployer)
- % réalisé changé de 0 à 100

Appliqué par commit [8f02fb31e7d11797fd5edaa01fc7b3b1127e7092](#).

#10 - 02 février 2018 18:05 - Thomas Noël

- % réalisé changé de 100 à 0

- * 8f02fb3 nanterre: améliore la recherche par nom (fixes #19092)
- * 31f1d2b nanterre: construire des index spécifiques pour les noms complets des individus (#19092)
- * 8684139 zoo_data: ajoute une nouvelle fonction SQL TextCat (#19092)
- * 6137d25 zoo_meta: ajoute un hook sur les classes AppConfig pour personnaliser la reconstruction des index (#19092)
- * efaafa5 zoo_meta: ajoute une méthode utilitaire pour créer des index trigram (#19092)
- * 0884695 zoo_meta: reconstruit les index via les signaux (#19092)

#11 - 02 février 2018 18:11 - Thomas Noël

J'ai quand même l'impression que 0001 relance l'indexation à chaque "migrate" ? (c'est à dire à chaque mise à jour de zoo...)?

#12 - 02 février 2018 18:12 - Thomas Noël

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

Déployé sur en recette Nanterre

#13 - 02 février 2018 18:26 - Thomas Noël

- Lié à Bug #21619: zoo nanterre: ne démarre plus parce que le migrate lance une reindexation full et TimeoutStartSec=30 ajouté

#14 - 02 février 2018 18:31 - Thomas Noël

- Lié à Bug #21620: crash sur détection de doublons ajouté

Fichiers

0005-nanterre-construire-des-index-sp-cifiques-pour-les-n.patch	1,28 ko	01 février 2018	Benjamin Dauvergne
0002-zoo_meta-ajoute-une-m-thode-utilitaire-pour-cr-er-de.patch	1,76 ko	01 février 2018	Benjamin Dauvergne
0003-zoo_meta-ajoute-un-hook-sur-les-classes-AppConfig-po.patch	1,14 ko	01 février 2018	Benjamin Dauvergne
0001-zoo_meta-reconstruit-les-index-via-les-signaux-19092.patch	1,91 ko	01 février 2018	Benjamin Dauvergne
0004-zoo_data-ajoute-une-nouvelle-fonction-SQL-TextCat-19.patch	987 octets	01 février 2018	Benjamin Dauvergne
0006-nanterre-am-liore-la-recherche-par-nom-fixes-19092.patch	15,4 ko	01 février 2018	Benjamin Dauvergne