

Combo - Development #43781

recherche : mieux prendre en compte les titres de page

08 juin 2020 17:28 - Paul Marillonnet

Statut:	Fermé	Début:	08 juin 2020
Priorité:	Normal	Echéance:	
Assigné à:		% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	Non
Patch proposé:	Oui		
Description			
Exemple sur la documentation PubliK, je souhaite accéder à une page de documentation dont je connais le titre, en l'occurrence "Utiliser les filtres". Je tape "utiliser les filtres" dans la cellule de recherche et la page ne ressort pas.			

Révisions associées

Révision 5356dbde - 26 juin 2020 14:31 - Paul Marillonnet

search: provide title precedence over content for indexed cells (#43781)

Historique

#1 - 08 juin 2020 17:31 - Paul Marillonnet

- Description mis à jour

#2 - 10 juin 2020 09:40 - Paul Marillonnet

Bon il semblerait que ce soit déjà le cas, le titre de l'élément IndexedCell (qui est le titre de la page de la cellule combo) fait partie du vecteur de recherche.

À mon avis, le système de pondération fait que les titres ne ressortent pas parmi les X premiers résultats, montrés à l'écran, lorsque je fais la recherche.

Si on est d'accord que le constat dans la description de ce ticket est problématique, je vais commencer par tester en local quelque chose comme

```
diff --git a/combo/apps/search/utils.py b/combo/apps/search/utils.py
index ff4942c7..2a43d861 100644
--- a/combo/apps/search/utils.py
+++ b/combo/apps/search/utils.py
@@ -113,7 +113,7 @@ def search_site(request, query, pages=None):
     if connection.vendor == 'postgresql':
         config = settings.POSTGRESQL_FTS_SEARCH_CONFIG
-        vector = SearchVector('title', config=config, weight='A') + SearchVector('indexed_text', config=config, weight='A')
+        vector = SearchVector('title', config=config, weight='A') + SearchVector('indexed_text', config=config, weight='B')
         query = SearchQuery(query, config=config)
         qs = IndexedCell.objects.annotate(rank=SearchRank(vector, query)).filter(rank__gte=0.3).order_by('-rank')
     else:
```

et voir si ça améliore la situation.

#3 - 10 juin 2020 09:46 - Frédéric Péters

Le constat "c'est dommage de ne pas trouver une page/certaines pages en tapant leur titre dans le moteur de recherche" est partagé, go.

#4 - 10 juin 2020 09:48 - Paul Marillonnet

- Sujet changé de recherche : indexer les titres de page ? à recherche : mieux prendre en compte les titres de page

#5 - 11 juin 2020 12:51 - Paul Marillonnet

- Fichier 0001-search-provide-title-precedence-over-content-for-ind.patch ajouté

- Statut changé de Nouveau à Solution proposée

- Patch proposed changé de Non à Oui

Rebasé sur [#43916](#).

#6 - 11 juin 2020 12:54 - Paul Marillonnet

- Fichier 0001-search-provide-title-precedence-over-content-for-ind.patch ajouté

Avec quelques vérifications supplémentaires qui me paraissent nécessaires dans le test.

#7 - 11 juin 2020 14:02 - Paul Marillonnet

- Statut changé de Solution proposée à En cours

Et effet de bord sur les autres tests de recherche car la différence de pondération entre A (=1) et B (=0.6) sur les vecteurs de recherche est trop forte. Je regarde si avec une graduation plus compacte en dessous de 1 on peut arranger l'affaire.

#8 - 11 juin 2020 14:38 - Paul Marillonnet

- Fichier 0001-search-provide-title-precedence-over-content-for-ind.patch ajouté

- Statut changé de En cours à Solution proposée

Paul Marillonnet a écrit :

Je regarde si avec une graduation plus compacte en dessous de 1 on peut arranger l'affaire.

En adaptant à cette diminution de poids le filtre sur le rang des résultats de recherche on y arrive.

#9 - 26 juin 2020 10:36 - Frédéric Péters

- Statut changé de Solution proposée à Solution validée

#10 - 26 juin 2020 14:37 - Paul Marillonnet

```
commit 5356dbde4eea33534c554c93c2757740a562fdb4
Author: Paul Marillonnet <pmarillonnet@entrouvert.com>
Date: Wed Jun 10 10:32:33 2020 +0200
```

```
search: provide title precedence over content for indexed cells (#43781)
```

#11 - 26 juin 2020 14:41 - Paul Marillonnet

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

#12 - 27 juin 2020 18:16 - Frédéric Péters

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

Fichiers

0001-search-provide-title-precedence-over-content-for-ind.patch	2,64 ko	11 juin 2020	Paul Marillonnet
0001-search-provide-title-precedence-over-content-for-ind.patch	2,71 ko	11 juin 2020	Paul Marillonnet
0001-search-provide-title-precedence-over-content-for-ind.patch	2,95 ko	11 juin 2020	Paul Marillonnet