

Amélioration du moteur de recherche "web"

28/04/2015 11:46 - Eric Seigne

| | | | |
|--|------------------------|----------------------|-----------------------------------|
| Statut: | Assigned - En cours | Début: | 28/04/2015 |
| Priorité: | Haute | Echéance: | |
| Assigné à: | Eric Seigne | % réalisé: | 80% |
| Catégorie: | data.abuledu.org (web) | Temps estimé: | 80.00 heures |
| Version cible: | | | |
| Description | | | |
| Pour permettre de faire les recherches sur les différents champs de scolomfr | | | |
| Demandes liées: | | | |
| Lié à abuledu-data, vidéos, mediatheque, serveur de REL - Bug #3604: Affichag... | | | Fixed - Corrigé 02/05/2014 |

Historique

#1 - 03/05/2015 02:47 - Eric Seigne

Big bug trouvé sur l'indexation lom/scolom: le fichier meta.xml (utilisé par xapian) n'était pas actualisé quand un fichier abe avait déjà été ouvert/utilisé par data avant !

-> rebuild en cours de tous les fichiers abe en prenant le lom.xml comme seule source d'information valable.

comme c'est > Qt 5.2 j'ai compilé le abuledu-cgi sur un autre serveur, trimbalé les 25550 ressources de abrier -> paco, corrige les fichiers sur paco et ensuite retour sur abrier ...

#2 - 03/05/2015 02:48 - Eric Seigne

J'ai oublié d'indiquer qu'on a déjà codé ça cette semaine:

<http://data.abuledu.org/wp/prototype-de-moteur-de-recherche/>

#3 - 05/05/2015 02:27 - Eric Seigne

La présentation des résultats se rapproche de ce qu'on souhaite. Les liens ne sont pas encore fonctionnels mais c'est la prochaine étape.

#4 - 17/05/2015 09:50 - Eric Seigne

Bon, ces trois derniers jours j'ai terminé l'implémentation du wordpress "client complet de l'entrepôt"

- Wordpress est de nouveau léger et "vide"
- supprimé les 25.000 articles et 15 Go d'images correspondantes
- supprimé le moteur de recherche "interne wordpress" et changé la page d'accueil (capture 001)
- Implémenté la présentation des mots clés & leur recherche
- Implémenté la page finale de visi d'une ressource (capture 002)

J'ai également repris le dev d'un prototype de moteur de recherche qui permettrait:

- de faire simplement des ajouts de conditions "NOT" dans la requête (avec les petites croix rouges affichées sur la capture 001)
- de faire tout ça en ajax
- de gérer les petites erreurs (exemple recherche lappins ou murene) car jusqu'à présent murene != murène != murènes

#5 - 17/05/2015 09:51 - Eric Seigne

- Fichier 20150517-data-001.jpg ajouté

- Fichier 20150517-data-002.jpg ajouté

#6 - 17/05/2015 09:54 - Eric Seigne

- % réalisé changé de 0 à 80

#7 - 18/05/2015 09:14 - Eric Seigne

Copier/coller du mail de compte rendu envoyé sur la liste dev@

1ere idée: les mots clés utilisés qu'une seule fois

La (longue) liste des mots clés utilisés qu'une seule fois, <http://data.abuledu.org/wp/nuage-des-mots-cles-utilises-quune-seule-fois/> ... ça vous donnera peut-être envie d'enrichir quelques ressources pour leur ajouter ce mot clé histoire que :-)

par contre si vous cliquez sur ce mot clé ou que vous faites une recherche spécifique sur ce mot clé vous aurez parfois plusieurs résultat, la faute à mon optimisation d'hier qui fait que lorsqu'on n'a qu'une seule réponse on essaye de chercher les singuliers et les suggestions de recherche, les mots proches ... ainsi la recherche sur le mot clé Amphiboles retourne bien la seule réponse mais par contre la recherche sur Arbres fruitiers est intéressante vu qu'on a une ressource qui a un mot clé mal orthographié (il manque un t) ... mais ne sera pas facile à trouver du fait de mon "correcteur orthographique" automatique ...

2eme idée: les mots clés les plus utilisés

C'est plus classique: <http://data.abuledu.org/wp/nuage-de-mots-cles/> mais ça permet de soulever la surprise de n'avoir que 566 ressources avec le mot clé Afrique ... c'est normal car par exemple <http://data.abuledu.org/wp/?LOM=19615> n'a pas ce mot clé. Je me lancerais donc bien dans le délire du pseudo thésaurus que je détaille juste après.

Et si j'avais du temps: un pseudo thésaurus

Là c'est jouer un peu à ce que fait google: si statistiquement le mot clé X est toujours associé avec Y alors je peux ajouter X à tous les documents qui parlent de Y et inversement :p ou tout au moins le suggérer ... pour les "relations statistiquement fortes" ainsi

<http://data.abuledu.org/wp/?LOM=19615> est à Abidjan donc en Afrique ... mais bon dans ce cas particulier l'ensemble des relations existantes sur data ne permettrait pas de le faire, alors que pour le Burkina Faso ça me semble plus jouable vu le nombre de documents qui ont le mot clé Burkina et Afrique.

Sans oublier qu'on a aussi l'excellent boulot de la BNF, par exemple on peut associer facilement Burkina Faso avec Afrique occidentale ainsi que la Côte d'Ivoire :

<http://data.abuledu.org/vocabRameau/vocab/index.php?tema=246927>

ou

<http://data.abuledu.org/vocabRameau/vocab/index.php?tema=158927>

#8 - 06/08/2015 02:45 - Eric Seigne

Trois bugs importants ont été constatés:

- la recherche sur cycle de l'eau ne donne rien
- l'eau se transforme en l'eau
- et ensuite quand on clique sur un mot clé magique comme abcd-cycle-eau aucune ressource n'apparaît

C'est corrigé :-)

Fichiers

| | | | |
|-----------------------|--------|------------|-------------|
| 20150517-data-001.jpg | 256 ko | 17/05/2015 | Eric Seigne |
| 20150517-data-002.jpg | 150 ko | 17/05/2015 | Eric Seigne |