Publié sur ***COAGUL*** (https://www.coagul.org/drupal)

Indexation et recherche de fichiers sous Linux

ven, 26/01/2007 - 20:29 — Tony

Fichiers et systèmes de fichiers [1]

But de ce document

Présenter les différents logiciels disponibles sous Linux permettant d'indexer et de rechercher des fichiers d'après leur nom ou d'après leur contenu.

Rappel des logiciels non libres disponibles sous Windows

Copernic Desktop

- http://www.copernic.com/fr/ [2]

Copernic est intégré à Windows et une barre d'outils est ajoutée dans Internet Explorer et dans Firefox.

Il permet l'indexation de nombreux fichiers :

- Thunderbird, OOo 1.x, OOo 2.x,...

Il est également possible d'indexer des fichiers sur le réseau.

Systèmes supportés : Windows 98/Me/NT4/2000/2000 Server/XP/2003 Server.

Google Desktop

- http://desktop.google.fr/fr/ [3]

L'installation de « Google Desktop », modifie la page d'accueil de Google et envoie des informations à Google. Google Desktop ne semble pas être capable d'indexer les fichiers OOo. L'indexation semble plus lente que pour Copernic. L'interface Web est moins conviviale que celle de Copernic.

Systèmes supportés : Win 2000, Win XP et Linux très récemment

Locate

Locate est le programme historique disponible sous Linux. Il n'indexe pas le contenu des fichiers mais permet déjà de trouver très facilement et rapidement un fichier d'après une partie de son nom.

L'indexation se fait avec la commande « **updatedb** »

La commande « **locate** » permet de recherche un fichier :

```
$ locate toto
```

Remarque : La recherche se fait par défaut dans le home de l'utilisateur lançant la commande et dans les dossiers systèmes.

kio-locate

Le paquet Debian « **kio-locate** » disponible sur le site officiel permet d'ajouter un kio-slave à KDE.

Pour l'utiliser, il faut par exemple saisir dans Konqueror (ou dans n'importe quelle applications KDE) l'URL « locate :/ » et d'ajouter le mot clé de recherche. Exemple :

```
locate:/toto
```

Lien : http://kde-apps.org/content/show.php ?content=17201 [4]

Strigi (Remplaçant de Kat qui est abandonné)

Strigi sera intégré à KDE 4. Il dispose déjà d'un applet pour KDE 3.5.

Installation :

```
# aptitude install strigi-client strigi-applet strigi-daemon
```

Strigi est utilisable en démon seul, mais il est possible d'installer un client et même un applet pour KDE.

Strigi n'est pas encore stable et personnellement je trouve Beagle beaucoup plus évolué.

Strigi repose sur le moteur d'indexation CLucene (Le portage en C++ du projet Lucene)

Beagle

L'indexation semble plus efficace que celle de Strigi. Cependant comme l'interface est faite pour Gnome (GTK), l'ouverture des fichiers trouvés ne fonctionne pas directement depuis « **beagle-search** » dans KDE (Mauvaise intégration dans KDE)

```
# aptitude install beagle
```

Techniquement Beagle utilise la version C# du moteur d'indexation Lucene écrit par la fondation apache. Il propose une interface client/serveur permettant d'effectuer des recherches à travers un client lourd (interface d-bus) **mais aussi un serveur web ou directement en exposant des web services.**

Beagle nécessite l'environnement Mono et GTK pour le client

Remarque : Il semblerait que pour faire fonctionner Beagle correctement il soit nécessaire d'activer les attributs étendus . Pour cela, il faut ajouter dans le fichier « fstab » et dans les options de la partition concernée le paramètre « **user_xattr** ». Exemple :

```
/dev/hda3 / ext3 defaults,errors=remount-ro,user_xattr 0 1
```

Ensuite, et pour éviter de redémarrer l'ordinateur, il faut remonter la partition pour prendre en compte ces modifications :

```
# mount -o remount /
```

Personnellement, j'ai testé Beagle sans activer ce paramètre et je n'ai vu aucune différence.

Lien : http://beagle-project.org/Main_Page [5]

Utiliser Beagle sous KDE avec Kerry

Le programme « **Kerry** » est simplement une interface graphique écrite en Qt pour Beagle. Elle utilise donc le même moteur et la même base de données d'indexation que « **beagle-search** » :

Après avoir installé Beagle, il suffit donc d'installer le programme « kerry » :

```
# aptitude install kerry
```

Kerry étant parfaitement intégré à KDE, il est possible d'ouvrir directement les fichiers depuis son interface de recherche.

Les fonctionnalités de l'interface de recherche de « **Kerry** » et « **beagle-search** » sont très proches, ce qui est normal étant

donné qu'ils reposent sur le même moteur. Mais j'ai quand même noté deux points présents dans « **Kerry** » mais pas dans « **beagle-search** » :

- Accéder au dossier contenant le fichier trouvé en cliquant sur le lien
- Limiter à la recherche aux fichiers modifiés pendant la journée, la semaine, le mois ou l'année

Le seul regret que j'ai avec « **Kerry** » est qu'il ne dispose pas d'un applet comme « **Strigi** » permettant de rechercher des fichiers directement depuis la barre des taches.

Remarques :

- Lors de la première indexation des fichiers, celle-ci se passe en essayant de ne pas surcharger le processeur. Mais en contre-partie, l'indexation de nombreux fichiers peut prendre plusieurs heures. L'onglet « Daemon statut » permet de voir si l'indexation est terminée et le nombre de fichiers indexés.
- Ensuite, tous les fichiers sont indexés à la volée et en temps réel dés l'enregistrement. La fenêtre de recherche des fichiers s'actualise automatiquement et instantanément à chaque enregistrement d'un fichier ce que je trouve assez impressionnant.

Problèmes rencontrés :

- J'ai désactivé l'indexation de Thunderbird car avec mes dossiers des mails de 100 Mo recevant des mails très régulièrement, l'indexation du fichier prenait 100% du processeur quasiment en permanence.
- En cas de modification des dossiers à indexer, il faut supprimer les anciens indexes manuellement pour ne pas retrouver les anciens fichiers et bien prendre en compte les nouveaux. Dans mon cas, j'ai supprimé complètement le dossier /.beagle
- Même en sélectionnant uniquement le « Backend » « Files », la recherche me retourne des données du « Backend » « Application »
- A priori, Beagle n'arrive pas à indexer tous les fichiers. Dans mon cas, il a indexé 2500 fichiers sur 3500. Il semblerait que cela soit dû aux dossiers et fichiers contenant des espaces ou des caractères spéciaux. Cela est peut-être lié au fait que les applications Gnome fonctionnent en UTF8 et que je suis configuré en ISO.

Kio-Beagle

Ce paquet ajoute un kio-salve à Konqueror accessible à l'URL «**beagle :/** »

Malheureusement le paquet n'est pas disponible sous Testing et le paquet Ubuntu n'est pas compatible.

Lien : http://linuxappfinder.com/package/kio-beagle [6]

htdig

htdig repose également sur le moteur CLucene, mais ce projet n'a pas évolué depuis 2004.

htdig permet d'indexer des sites web.

Le programme « o3read » permet d'indexer des fichiers OOo

Autres liens

- http://www.vandenoever.info/software/strigi/ [7]
- http://beagle-project.org/Main_Page [5]

Historique des modifications

Version	Date	Commentaire
0.1	30/08/06	Création par Tony GALMICHE
0.4	30/11/06	Mise en ligne
0.5	14/08/07	Ajout « Linux » pour « Google Desktop »

Licence Creative Commons by-sa 3.

URL source: https://www.coagul.org/drupal/publication/indexation-et-recherche-fichiers-sous-linux

Liens:
[1] https://www.coagul.org/drupal/rubrique/fichiers-et-systemes-fichiers
[2] http://www.copernic.com/fr/
[3] http://desktop.google.fr/fr/
[4] http://kde-apps.org/content/show.php?content=17201
[5] http://beagle-project.org/Main_Page
[6] http://linuxappfinder.com/package/kio-beagle
[7] http://www.vandenoever.info/software/strigi/