fdupes et la recherche de doublons

fdupes est utilitaire en ligne de commande qui permet de retrouver très rapidement les fichiers en doublon présents sur vos espaces de stockage. Pour repérer les doublons, il ne se base pas sur le nom des fichiers mais les identifie en comparant **la signature MD5 des fichiers suivie d'une comparaison octet à octet**. Cet utilitaire, écrit en C, sous licence MIT, possède de très nombreuses options afin de répertorier, supprimer et remplacer les fichiers.

Voici la chaîne de comparaison utilisée

```
1/ Comparaison de taille
2/ Comparaison de signature MD5 partielle
3/ Comparaison de signature MD5 complète
4/ Comparaison octet à octet.
```

Installation

Sous base DEBIAN

```
# apt-get install fdupes
```

Sous base UBUNTU

```
$ sudo apt-get install fdupes
```

Sous freeBSD, depuis les repos

```
 # pkg install fdupes
```

Sous freeBSD, depuis les ports

```
# cd /usr/ports/sysutils/fdupes/
# make install clean
```

Utilisation

Syntaxe :

```
fdupes <option> <répertoire>
```

Soit un répertoire, **dupli-test** contenant 3 fichiers distincts de nom et de contenu.

```
$ ls test-dupli/
```

```
aze1.txt  aze3.txt  aze.txt
```

Utilisons un GUI et copions les, via **ctrl +C / ctrl + V** :

```
$ ls test-dupli/
'aze1 (copie 1).txt'  'aze3 (copie 1).txt'  'aze (copie 1).txt'
'aze1 (copie 2).txt'  'aze3 (copie 2).txt'  'aze (copie 2).txt'
'aze1 (copie 3).txt'  'aze3 (copie 3).txt'  'aze (copie 3).txt'
'aze1 (copie 4).txt'  'aze3 (copie 4).txt'  'aze (copie 4).txt'
'aze1 (copie 5).txt'  'aze3 (copie 5).txt'  'aze (copie 5).txt'
'aze1 (copie 6).txt'  'aze3 (copie 6).txt'  'aze (copie 6).txt'
 aze1.txt              aze3.txt              aze.txt
```

Pour rechercher les doublons, en être informé, sans aucune autre action, pas besoin d'option :

```
$ fdupes  test-dupli/
test-dupli/aze3.txt
test-dupli/aze3 (copie 1).txt
test-dupli/aze3 (copie 3).txt
test-dupli/aze3 (copie 2).txt
test-dupli/aze3 (copie 6).txt
test-dupli/aze3 (copie 5).txt
test-dupli/aze3 (copie 4).txt

test-dupli/aze.txt
test-dupli/aze (copie 1).txt
test-dupli/aze (copie 3).txt
test-dupli/aze (copie 2).txt
test-dupli/aze (copie 5).txt
test-dupli/aze (copie 4).txt
test-dupli/aze (copie 6).txt

test-dupli/aze1.txt
test-dupli/aze1 (copie 1).txt
test-dupli/aze1 (copie 3).txt
test-dupli/aze1 (copie 2).txt
test-dupli/aze1 (copie 4).txt
test-dupli/aze1 (copie 6).txt
test-dupli/aze1 (copie 5).txt
```

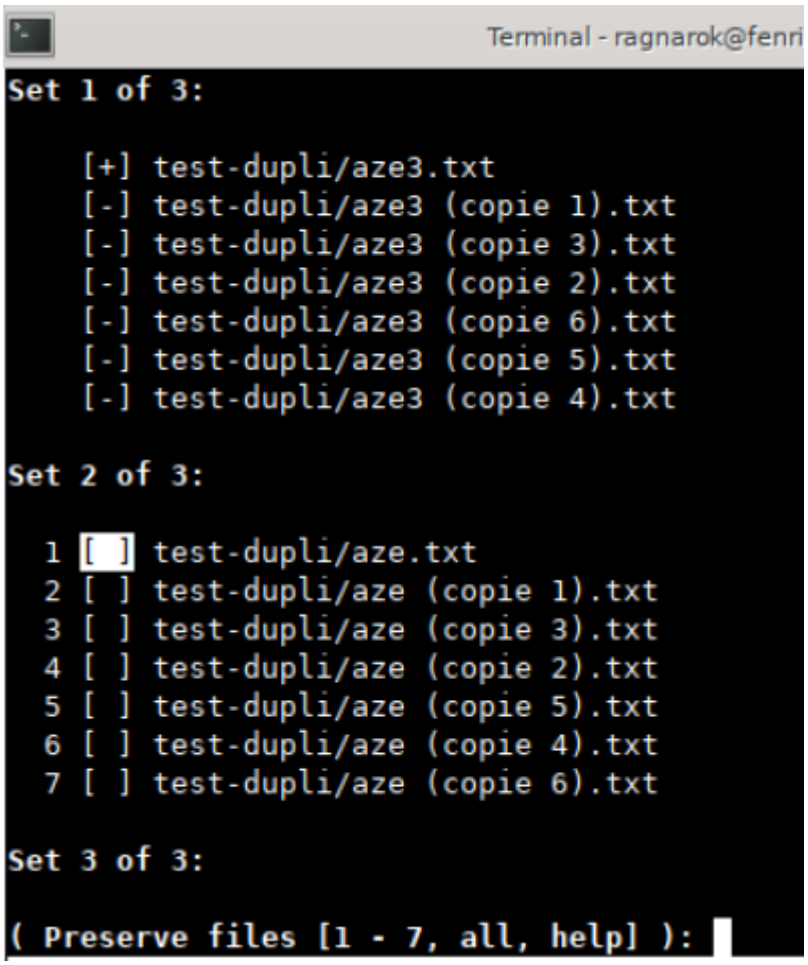
Pour rechercher les doublons, en être informé, et connaitre l'espace disque qu'ils utilisent, **utiliser l'option -S (Size)** :

```
$ fdupes -S test-dupli/
13 bytes each:
test-dupli/aze3.txt
test-dupli/aze3 (copie 1).txt
....
test-dupli/aze3 (copie 4).txt

169 bytes each:
```

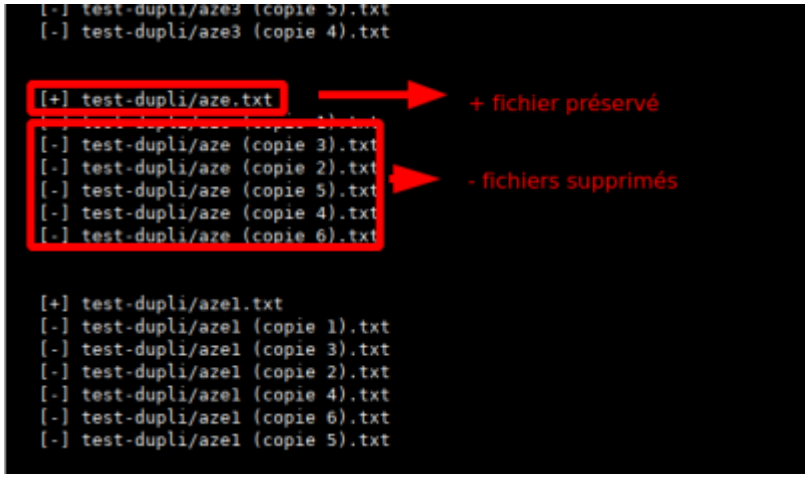
```
test-dupli/aze.txt
test-dupli/aze (copie 1).txt
....
test-dupli/aze (copie 6).txt

49 bytes each:
test-dupli/aze1.txt
test-dupli/aze1 (copie 1).txt
....
test-dupli/aze1 (copie 5).txt
```

__ Pour supprimer les doublons, **utiliser l'option -d (delete)** :__

Pour chaque doublon, une liste sera proposée, il faudra indiquer le fichier à préserver et les autres seront marqués à supprimer.



Pour supprimer les doublons, sans invite prompt (le nom le plus simple sera alors préserver), utiliser les options -d et -N (Non-prompt) La suppression sera alors immédiate.



__ Pour faire de de même de façon récursive, **ajouter aux options précédentes simplement l'option -r** :__

```
$ fdupes -rdN .
```

Donnera une suppression récursive des doublons du répertoire courant (**.**)

```
$ fdupes -rS /home/$USER
```

Donnera la liste des doublons et leurs occupations disques pour le /home de l'utilisateur courant.

Pour sauvegader cette liste dans un fichier, utiliser une redirection

```
$ fdupes -rS /home/$USER > /home/$USER/liste-de-doublons
```

A savoir, 2 fichiers au même contenu mais avec des noms différents seront considérés comme doublon, même avec des extensions différentes.

```
$ cat test-dupli/aze.txt test-dupli/test.sh
Test de doubon. Test de doubon. Test de doubon. Test de doubon. Test de
doubon.
Test de doubon. Test de doubon. Test de doubon. Test de doubon. Test de
doubon.
```

On voit bien que les 2 fichiers ont exactement le même contenu.

```
$ fdupes  -rS test-dupli/
81 bytes each:
test-dupli/aze.txt
test-dupli/test.sh
```

fdupes les analyse bien comme doublon.

Pour ne pas prendre en compte les fichiers cachés, **utiliser l'option -A**.

Pour des plus amples informations

```
$ man fdupes
```

```
$ fdupes --help
$ fdupes -h
```

From:
https://cbiot.fr/dokuwiki/ - **Cyrille BIOT**

Permanent link:
https://cbiot.fr/dokuwiki/fdupes

Last update: **2020/09/13 19:13**