

upandclear.org

rdfind : trouver des fichiers dupliqués. Ou “le Plex DupeFinder” du pauvre

Par : Aerya

7-9 minutes

[home](#) > [Articles](#) > rdfind : trouver des fichiers dupliqués. Ou “le Plex DupeFinder” du pauvre

évidemment on peut remplacer “pauvre” par geek ou noob ou celui qui veut faire simple.

Pour ceux qui ne connaissent pas, Plex DupeFinder est un script Python dédié à Plex (Captain Obvious!) et qui permet de lister/supprimer les fichiers en double. Parce que tout le monde sait que ça peut vite devenir le boxon dans un espace de stockage entre l’autoDL, les RSS, les clics sur divers sites, le pote qui ajoute un fichier, la copine etc. Si vous voulez découvrir ça, le script est [ici](#) et le “howto” [là](#) (ouais je sais, pas logique).

Si vous avez cliqué sur le second lien vous avez du vite piger que c’est plutôt pénible à mettre en place. Rassurez-vous, Linux propose de multiples solutions. Moins hypées mais plus simples. J’avais déjà parlé de [FDupes](#), dans cet article ce sera [rdfind](#).

Pour l'installer sur Ubuntu/Debian :

Pour le lancer il suffit de lui indiquer quel dossier traiter.

Selon le nombre de fichiers ça peut prendre beaucoup de temps... L'option “-dryrun true” permet de lancer rdfind sans qu'il n'effectue aucune action sur les fichiers concernés, il se contentera de trouver les fichiers dupliqués et de faire un export (une liste) dans un fichier texte.

1. rdfind -dryrun true mnt/unionfs/Films

Il va partir du nombre de total de fichiers et retirer ceux de 0 octet (oui, j'assume le bazar sur mon serveur), éliminer rapidement ceux dont les 1er/dernier bits sont différents et enfin comparer les hashes ([MD5](#) / [SHA1](#) / [SHA256](#)) des fichiers restants :

1. (DRYRUN MODE) Now scanning "mnt/unionfs/Films", found 66 files.
2. (DRYRUN MODE) Now have 66 files in total.
3. (DRYRUN MODE) Removed 0 files due to nonunique device and inode.
4. (DRYRUN MODE) Now removing files with zero size from list...removed 6 files
5. (DRYRUN MODE) Total size is xxx bytes or xx Tib
6. (DRYRUN MODE) Now sorting on size:removed 10 files due to unique sizes from list.50 files left.
7. (DRYRUN MODE) Now eliminating candidates based on first bytes:removed 2 files from list.48 files left.
8. (DRYRUN MODE) Now eliminating candidates based on

last bytes:removed 4 files from list.46 files left.

9. (DRYRUN MODE) Now eliminating candidates based on md5 checksum:removed 32 files from list.14 files left.
10. (DRYRUN MODE) It seems like you have 14 files that are not unique
11. (DRYRUN MODE) Totally, xx Gib can be reduced.
12. (DRYRUN MODE) Now making results file results.txt

J'ai évidemment modifié les valeurs. Mais on voit quand même la logique. Comme il l'indique à la fin, il a listé dans le fichier results.txt les 14 fichiers dupliqués qu'il a trouvés et qui, une fois supprimés, me feraient économiser quelques Gigas.

1. # Automatically generated
2. # duptype id depth size device inode priority name
3. DUPTYPE_FIRST_OCCURRENCE 3204 1 312 1048816
8550 1 mnt/unionfs/Films
/abc.2013.MULTi.1080p.BluRay.x264-
ZEST/abc.2013.MULTi.1080p.BluRay.x264-ZEST.nfo
4. DUPTYPE_WITHIN_SAME_TREE -3204 1 312 1048816
8553 1 mnt/unionfs/Films
/abc.2013.MULTi.1080p.BluRay.x264-
ZEST/abc.2013.multi.1080p.bluray.x264-zest.nfo
5. DUPTYPE_FIRST_OCCURRENCE 1845 0 1851051999
1048816 6742 1 mnt/unionfs/Films
/def.2018.FRENCH.1080p.NF.WEB-DL.DD5.1.H264-
PHI.mkv

6. DUPTYPE_WITHIN_SAME_TREE -1845 1 1851051999
1048816 6741 1 mnt/unionfs/Films
/def.2018.FRENCH.1080p.NF.WEB-DL.DD5.1.H264-
PHI/def.2018.FRENCH.1080p.NF.WEB-DL.DD5.1.H264-
PHI.mkv
7. DUPTYPE_FIRST_OCCURRENCE 2019 0 1922499993
1048816 6981 1 mnt/unionfs/Films
/ghi.2018.FRENCH.1080p.WEB-DL.x264.mkv
8. DUPTYPE_WITHIN_SAME_TREE -2019 1 1922499993
1048816 6980 1 mnt/unionfs/Films
/ghi.2018.FRENCH.1080p.WEB-DL.x264
/ghi.2018.FRENCH.1080p.WEB-DL.x264.mkv
9. DUPTYPE_FIRST_OCCURRENCE 3135 0 2060451438
1048816 8449 1 mnt/unionfs/Films/jkl 1080p x265 HEVC
AC3 Multi VFQ-VO.mkv
10. DUPTYPE_WITHIN_SAME_TREE -3135 1 2060451438
1048816 8451 1 mnt/unionfs/Films/jkl/jkl 1080p x265 HEVC
AC3 Multi VFQ-VO.mkv
11. DUPTYPE_FIRST_OCCURRENCE 485 0 2915972590
1048816 4933 1 mnt/unionfs/Films/mno MULTi [1080p]
BluRay x264.mkv
12. DUPTYPE_WITHIN_SAME_TREE -485 1 2915972590
1048816 4932 1 mnt/unionfs/Films/mno MULTi [1080p]
BluRay x264/mno MULTi [1080p] BluRay x264.mkv
13. DUPTYPE_FIRST_OCCURRENCE 1654 0 3062109337
1048816 6495 1 mnt/unionfs/Films
/pqr.2018.MULTi.1080p.NF.WEB-DL.DD5.1.x264-

LIBERTAD.mkv

14. DUPTYPE_WITHIN_SAME_TREE -1654 1 3062109337
1048816 6493 1 mnt/unionfs/Films
/pqr.2018.MULTi.1080p.NF.WEB-DL.DD5.1.x264-LIBERTAD
/pqr.2018.MULTi.1080p.NF.WEB-DL.DD5.1.x264-
LIBERTAD.mkv
15. DUPTYPE_FIRST_OCCURRENCE 2473 0 3109742029
1048816 7555 1 mnt/unionfs/Films
/stu.2018.MULTi.1080p.WEB.H264-GAZOAL.mkv
16. DUPTYPE_WITHIN_SAME_TREE -2473 1 3109742029
1048816 7554 1 mnt/unionfs/Films
/stu.2018.MULTi.1080p.WEB.H264-GAZOAL
/stu.2018.MULTi.1080p.WEB.H264-GAZOAL.mkv
17. DUPTYPE_FIRST_OCCURRENCE 2172 0 3313404092
1048816 7164 1 mnt/unionfs/Films
/vwx.2018.FRENCH.1080p.WEB.H264-DOLL4R.mkv
18. DUPTYPE_WITHIN_SAME_TREE -2172 1 3313404092
1048816 7163 1 mnt/unionfs/Films/vwx
2018/vwx.2018.FRENCH.1080p.WEB.H264-DOLL4R.mkv
19. DUPTYPE_FIRST_OCCURRENCE 4038 0 3323052032
1048816 9666 1 mnt/unionfs/Films/Yyz.2011.1080p.
Multi.AC3-DTS.x264-Mexl.mkv
20. DUPTYPE_WITHIN_SAME_TREE -4038 1 3323052032
1048816 9665 1 mnt/unionfs/Films/yz.2011.1080p.
Multi.AC3-DTS.x264-Mexl/yz.2011.1080p. Multi.AC3-
DTS.x264-Mexl.mkv
21. DUPTYPE_FIRST_OCCURRENCE 2505 0 3509676342

- 1048816 7600 1 mnt/unionfs/Films
/123.2018.FRENCH.1080p.WEB.H264-PiCKLES.mkv
22. DUPTYPE_WITHIN_SAME_TREE -2505 1 3509676342
1048816 7599 1 mnt/unionfs/Films
/123.2018.FRENCH.1080p.WEB.H264-PiCKLES
/123.2018.FRENCH.1080p.WEB.H264-PiCKLES.mkv
23. DUPTYPE_FIRST_OCCURRENCE 2254 0 3527773871
1048816 7264 1 mnt/unionfs/Films
/456.2017.MULTi.1080p.NF.WEB-DL.DD5.1.x264-
ARK01.mkv
24. DUPTYPE_WITHIN_SAME_TREE -2254 1 3527773871
1048816 7263 1 mnt/unionfs/Films/456
2017/456.2017.MULTi.1080p.NF.WEB-DL.DD5.1.x264-
ARK01.mkv
25. DUPTYPE_FIRST_OCCURRENCE 1900 0 3552573283
1048816 6820 1 mnt/unionfs/Films/789 - 1080p FR x264
ac3 .mkv
26. DUPTYPE_WITHIN_SAME_TREE -1900 1 3552573283
1048816 6819 1 mnt/unionfs/Films/789 - 1080p FR x264
ac3 /789 - 1080p FR x264 ac3.mkv
27. DUPTYPE_FIRST_OCCURRENCE 3133 0 3648407391
1048816 8448 1 mnt/unionfs/Films/321 MULTi VF [1080p]
HDTV x264.mkv
28. DUPTYPE_WITHIN_SAME_TREE -3133 1 3648407391
1048816 8447 1 mnt/unionfs/Films/321 MULTi VF [1080p]
HDTV x264/321 MULTi VF [1080p] HDTV x264.mkv
29. DUPTYPE_FIRST_OCCURRENCE 1233 0 3710873036

1048816 5926 1 mnt/unionfs/Films/654 2018

TRUEFRENCH.1080p WEBRip.H264-KwAnGn0fX.mkv

30. DUPTYPE_WITHIN_SAME_TREE -1233 1 3710873036

1048816 5925 1 mnt/unionfs/Films/654 2018

TRUEFRENCH.1080p WEBRip.H264-KwAnGn0fX/654

2018 TRUEFRENCH.1080p WEBRip.H264-

KwAnGn0fX.mkv

31. # end of file

Et maintenant on peut faire le ménage en relançant rdfind mais cette fois-ci avec la commande de suppression (des fichiers de 0 octet et) des dupliqués :

1. rdfind -deleteduplicates true -ignoreempty false mnt/unionfs /Films

On peut aussi utiliser cette commande directement sans passer par l'étape dryrun. Rdfind ne peut pas se tromper dans ses résultats/actions. Je trouve cependant intéressant de décomposer l'opération qui de toute manière ne se fait que très rarement.

1. Now scanning "mnt/unionfs/Films", found 66 files.

2. [...]

3. Totally, xx Gib can be reduced.

4. Now making results file results.txt

5. Now deleting duplicates:

6. Deleted 14 files.

(28 vues)