

How to find duplicate files BY CONTENT!!

Today in IRC suseROCKS needed to find all duplicate files in a directory by their content, not by their file name, so we whipped up this fancy little 1 liner bash script to do the trick:

```
find . -type f -exec md5sum '{}' \; | sort | awk 'dup[$1]++{print $2}'
```

EDIT:

As Andreas suggested, using xargs instead of -exec is much faster, here is the updated command:

```
find . -type f -print0 | xargs -0 md5sum | sort | awk 'dup[$1]++{print $2}'
```

Revision #3

Created 2026-04-01 17:11:22 CEST by Philip

Updated 2026-04-13 19:24:54 CEST by Philip