

## NAME

findeqf – find equal files in a directory trees

## SYNOPSIS

```
findefq [ -bdata_base -l -v -Ssort_sequence ] [ directory ... ]
groupeqf data_base1 data_base2
```

## DESCRIPTION

*Findeqf* finds equal (duplicate) files in the indicated directory tree(s), and writes the name pairs to standard output. A missing directory name is interpreted as the working directory. The name pairs are reported as

$$name1 = ||| = name2$$

where one = indicates that the contents of *name1* are equal to those of *name2* ; two =s indicate that the modification times of both files are equal too; and three =s indicate that in addition their base names are equal. A usual call is

findeqf

which finds and reports duplicate files in the working directory and its subdirectories.

The `-b`-option causes *findgef* to store and/or retrieve an abstract of the information (file length, hash value, and file name) in or from the indicated data base. The data base is created if it does not exist. If it does exist, information from it is retrieved and used in the comparisons. This allows comparison to files that have been offloaded. Matches found from databases are reported with a question mark, for example:

$$subdir/foo \stackrel{?}{=} qwert/bar$$

where *subdir/foo* is from the present file system, and *quert/bar* is from the data base.

The `-l`-option restricts the search to the local files in the directory.

The `-v`-option produces a report in a more narrative style.

*Finddef* tries to report the duplicate files in the order *copy = original* , but it is difficult to tell which is which. Four criteria can be used:

- The file with the most recent modification time is the copy; this is criterion  $t$ .
- The file in the deepest directory is the copy; this is criterion  $d$ .
- The file with the longest base name is the copy; this is criterion  $l$ .
- The file with the alphabetically later name is the copy; this is criterion  $a$ .

In default mode *finddef* applies the criteria in the above order, but the **-S**-option can be used to select a different set and a different order of criterion applications. For example, *-Sdl* will first test for directory depth and then for file name length. When the tests from a **-S**-option do not yield a decision, the file with the alphabetically later name is assumed to be the copy. More in particular, the single option *-S* achieves that effect.

*Groupegf* accepts two database files as produced by earlier calls of *findegf* and produces a list of pairs of directories, one from each database and one from the second, indicating the total size of the files the two directories have in common; the names of the individual files are not shown and need not match.

**AUTHOR.**

Dick Grune, Vrije Universiteit, Amsterdam.