

NAME

comtar – Compare tar.gz archive files.

SYNOPSIS

`${HOME}/bin/comtar A/dir.tar.gz B/dir.tar.gz`

MOTIVATION

The most straightforward way to compare two tar.gz archives is to unpack them both and compare the resulting directories and their subdirectories with `/usr/bin/diff -r`. The `-q` option can be added to report only the names of differing files, which is the information most often of interest in comparing archives. Unfortunately, `.pdf` files representing the same document (and perhaps other binary files that are equivalent to each other in function) are reported to be different because of the metadata they contain; the `-a` option of `diff` can be used to print the differing metadata, but there is no way to make `diff` recognize and ignore it.

DESCRIPTION

Instead this script lists the members of the two archives and uses the `diff` program to compare the lists. If the members differ in names or in number, the archives cannot be the same; if corresponding members all have the same names and byte counts, the archives are assumed to be the same. Of course files having the same size can differ, so this heuristic is appropriate only in a context where it can be assumed that all interesting changes to the content of a file will also change its size by at least 1 byte.

If a file has different sizes in the two archives, this script extracts the file from each. If the name has the extension `.pdf`, the extracted files are compared using the `compdf` script; otherwise the files are compared using `/usr/bin/diff -b`, which ignores trailing blanks. The verdict of this comparison is reported.

UNITS

The script writes messages on standard-out.

SEE ALSO

`diff`, which this script runs
`compdf`, which this script invokes

DIAGNOSTICS

These are the possible return codes.

- 0 no differences were found between the two archives
- 1 at least one file was found to differ
- 2 the members differ in names
- 3 the members differ in number
- 4 one or both archives contain too many members

BUGS

Except for `.pdf` files, binary files containing metadata are reported to be different when they might be functionally equivalent.

AUTHOR

Michael Kupferschmid

EXAMPLE

```
unix[1] comtar Website.old/flash.tar.gz Website/flash.tar.gz
member names match but some byte counts differ
```

```
flash4.pdf      compare equal
flash12.pdf     compare equal
flash14.pdf     compare equal
flash28.pdf     compare equal
flash6.pdf      compare equal
```

```
unix[2] comtar Website.old/hebpgm.tar.gz Website/hebpgm.tar.gz
member names match but some byte counts differ
```

```
heblex.f        differ
```

```
unix[3] comtar Website.old/impsrc.tar.gz Website/impsrc.tar.gz
member names and sizes match
```

```
unix[4]
```