
COMPAREDIRECTORIES

By: Larry Masinter and Ron Kaplan

This document edited on

December 2, 1987

December 28, 1998 (Ron Kaplan)

April 7, 2018 (Ron Kaplan)

Rewritten August 25, 2020 (Ron Kaplan)

COMPAREDIRECTORIES compares the contents of two directories, identifying files according to their creation dates and lengths. It is called using the function

`(COMPAREDIRECTORIES DIR1 DIR2 SELECT FILEPATTERNS EXTENSIONSTOAVOID
USEDIRECTORYDATES OUTPUTFILE ALLVERSIONS SHORTDIRNAMES)` [Function]

Compares the creation dates of files with matching names in the lists that CDFILES returns for DIR1 and DIR2. Collects or prints CDENTRIES for those files that meet the SELECT criteria. May also collect or print entries for relevant files that exist in DIR1 or DIR2 but not both.

SELECT specifies which the match/mismatch criteria for filtering the output. If SELECT is or contains

AFTER or >: select entries where file1 has a later date than file2

BEFORE or <: select entries where file1 has an earlier date than file2

SAMEDATE or =: select entries where file1 and file2 have the same date

-*: exclude entries where file1 does not exist

*-: exclude entries where file2 does not exist

~=: exclude entries where file1 and file2 are byte-equivalent

SELECT = NIL is equivalent to (< > -* *-). Excludes files with matching dates, a useful default for identifying files that may require further attention.

SELECT = T is equivalent to (= < > -* *-). Includes all files for processing by other functions or later filtering by CDSUBSET (below).

SELECT may also contain the token AUTHOR to indicate that authors should be provided in the printed output (see CDPRINT below).

Unless USEDIRECTORYDATES, the FILECREATED date is used for the date comparison of Lisp source and compiled files, otherwise the file-system CREATIONDATE is used.

If OUTPUTFILE=NIL, then a list of the form (Parameters . entries) is returned. Parameters is a list (DIR1 DIR2 SELECT DATE) that records the parameters of the comparison. Entries contains one entry for each of the file-comparisons that meets the SELECT criteria. Each entry is a CDENTRY record with fields

(matchname info1 daterel info2 equiv)

where matchname is the name.extension shared by the two files, and each file info is either NIL (for nonexistent files) or a CDINFO record with fields

(FULLNAME DATE LENGTH AUTHOR TYPE EOL)

TYPE is SOURCE for Lisp source (filecreated) files, COMPILED for Lisp compiled files, otherwise the PRINTFILETYPE (TEXT, TEDIT...) or NIL. EOL is CR, LF, CRLF, or NIL.

When both files exist, the date relation is one of <, =, or >. Otherwise, the date relation is * if only one file exists. EQUIV is EQUIVALENT for files with different dates but exactly the same bytes, otherwise NIL.

If OUTPUTFILE is not NIL, then it is a filename or open stream on which selected entries will be printed (T for the terminal) by CDPRINT.

SHORTDIRNAMES can be a pair of strings to be used instead of the full directory names when header information is printed.

COMPAREDIRECTORIES sets the variable LASTCDENTRIES is set to the selected entries. This is used by the functions below if their CDENTRIES is NIL.

(CDFILES DIR FILEPATTERNS EXTENSIONSTOAVOID ALLVERSIONS DEPTH) [Function]

Returns a list of full filenames for files in directory DIR (NIL=T=the connected directory) that match the other file-name filtering criteria. Files are excluded if:

Their name does not match a pattern in FILEPATTERNS (NIL = *). Dotted files are excluded unless FILEPATTERNS includes .* and files in subdirectories are excluded if the number of subdirectories exceeds DEPTH (below).

Their extension is in the list EXTENSIONSTOAVOID (* excludes all extensions).

They are not the highest version unless ALLVERSIONS=T.

DEPTH controls the depth of subdirectory exploration. T means all levels, NIL means no subdirectories. Otherwise the maximum number of ">" characters below the starting DIR in the fullname of files.

(CDFILES) produces all the newest, undotted files in the immediate connected directory.

(CDPRINT CDENTRIES FILE PRINTAUTHOR) [Function]

Prints CDENTRIES on FILE, with one line for each entry. The line for each entry is of the form

FILE1 (AUTHOR) SIZE DATE relation DATE FILE2 (AUTHOR) SIZE

For example

```
ACE.;1 (Joe) 4035 2-May-1985 18:03:54 < 30-Sep-1985 11:14:48 ACE.;3 (Sam) 5096.
```

The line for byte-equivalent files is prefixed with ==. If the files are equivalent except for a difference in end-of-line conventions, the equivalence prefix will indicate the convention for each file (C for CR, L, for LF, 2 for CRLF). Thus C2 indicates that the files are equivalent except that file1 marks line ends with CR and file2 with CRLF.

Note that because of the setting of LASTCDENTRIES, evaluating (CDPRINT) after COMPAREDIRECTORIES prints the results of the last comparison.

For conciseness, authors are included only if PRINTAUTHOR or if AUTHOR is included in the SELECT parameter of CDENTRIES. Also, the redundant file-name hosts/directories are not printed.

(CDMAP CDENTRIES FN) [Function]

(CDSUBSET CDENTRIES FN) [Function]

CDMAP applies FN to each CDENTRY in CDENTRIES. CDSUBSET applies FN and also returns the subset of CDENTRIES for which FN is non-NIL and preserves in the value the parameters of CDENTRIES. For convenience, at each invocation the variables MATCHNAME INFO1 DATEREL INFO2 and EQUIV are bound to the corresponding fields and can be used freely by FN.

USEFUL UTILITIES

(FIX-DIRECTORY-DATES FILES) [Function]

For every file included in or specified by FILES, if it is a Lisp source or compiled whose directory creation date is more than 30 seconds later than its internal filecreated date (presumably because of copying), then its directory date is reset to match the internal date. FILES can be a list of file names or a pattern interpretable by FILDIR. Returns a list of files whose dates have been changed.

(FIX-EQUIV-DATES CDENTRIES) [Function]

If there is an entry in CDENTRIES whose files are EQUIVALENT but with different directory creation dates, the directory date of the file with the later date (presumably a copy) is reset to match the date of the earlier file. In the end all equivalent files will have the same (earliest) date. Returns a list of files whose dates have been changed.

(COPY-MISSING-FILES CDENTRIES TARGET MATCHNAMES) [Function]

Target is 1 or 2, indicating the direction of potential copies. If an entry with a source file but no target file has a matchname in MATCHNAMES, the source file is copied to the target directory. All target-absent files are copied if MATCHNAMES is NIL. Source properties (including version number) are preserved in the target.

(COPY-COMPARED-FILES CDENTRIES TARGET MATCHNAMES) [Function]

Target is 1 or 2, indicating the direction of potential copies. If an entry with both source and target files has a matchname in MATCHNAMES, the source file is copied to a new version of the target file. All files are copied if MATCHNAMES is NIL.

(COMPARE-ENTRY-SOURCE-FILES CDENTRY LISTSTREAM EXAMINE DW?) [Function]

This is a simple wrapper for calling COMPARERESOURCES if the CDENTRY files are Lisp source files. The function (CDENTRY MATCHNAME CDENTRIES) is useful for extracting a particular entry, with CDENTRIES defaulting to LASTCDENTRIES.

(COMPILED-ON-SAME-SOURCE CDENTRIES) [Function]

Returns the subset of entries with Lisp compiled files (dfasl or lcom) that are compiled on the same source, according to SOURCE-FOR-COMPILED-P below. Presumably one should be removed to avoid confusion.

(FIND-SOURCE-FILES CFILES SDIRS DFASLMARGIN) [Function]

Returns (CFILE . SFILES) pairs where CFILE is a Lisp compiled file in CFILES and SFILES is list of files in SDIRS that CFILE was compiled on according to SOURCE-FOR-COMPILED-P. This suggests that at least one of SFILES should be copied to CFILE's location (or vice versa).

(FIND-COMPILED-FILES SFILES CDIRS DFASLMARGIN) [Function]

Returns (SFILE . CFILES) pairs where SFILE is a Lisp source file in SFILES and CFILES are files in CDIRS that are compiled on SFILE according to SOURCE-FOR-COMPILED-P. This suggests that at least one of CFILES should be copied to SFILE's location.

(FIND-UNCOMPILED-FILES FILES DFASLMARGIN COMPILEXTS) [Function]

Returns a list of elements each of which corresponds to a source file in FILES for which no appropriate compiled file can be found. An appropriate compiled file is a file in the same location with extension in COMPILEEXTS (defaulting to *COMPILED-EXTENSIONS*) that satisfies SOURCE-FOR-COMPILED-P. Each element is a list of the form

(sourcefile . cfiles)

cfiles contains compiled files that were compiled on a different version of sourcefile, NIL if no such files exist. Each cfile item is a pair (cfile timediff) where timediff is the time difference (in minutes) between the creation date of the compiled-file's source and the creation date of sourcefile (positive if the cfile was compiled later, as should be the case). FILES can be an explicit list of files, or a file specification interpretable by FILDIR; in that case only the newest source-file versions are processed.

(FIND-UNSOURCED-FILES CFILES DFASLMARGIN COMPILEXTS) [Function]

Returns the subset of the compiled files specified by CFILES for which a corresponding source file according to SOURCE-FOR-COMPILED-P cannot be found in the same directory. CFILES can be a list of files or a pattern that FILDIR can interpret. COMPILEEXTS can be one or more explicit compile-file extensions, defaulting to *COMPILED-EXTENSIONS*.

(SOURCE-FOR-COMPILED-P SOURCE COMPILED DFASLMARGIN) [Function]

Returns T if it can confirm that Lisp COMPILED file was compiled on Lisp SOURCE file. SOURCE and COMPILED can be provided as CREATED-AS values, to avoid repetitive computation. This compares the information in the filecreated expressions, original file names and original dates, and not the current directory names and dates.

It appears that the times in DFASL files may differ from the filecreated source dates by a few minutes. The DFASLMARGIN can be provided to loosen up the date matching criterion. DFASLMARGIN is a

pair (max min) and a DFASL COMPILED is deemed to be compiled on SOURCE if the compiled's source date is no more than max and no less than min minutes after the source date. A negative min allows for the possibility that the compiled-source date is earlier than the candidate source date. DFASLMARGIN defaults to (20 0). A single positive number x is coerced to (x 0). A single negative number is coerced to (-x x) (compiled file is no more than x minutes later or earlier). T is infinity in either direction. Examples:

(T 0): COMPILED compiled on source later than SOURCE
 (0 T): COMPILED compiled on source earlier than SOURCE (odd)
 12: COMPILED compiled on source later than SOURCE by no more than 12 minutes
 -12: COMPILED compiled on source 12 minutes before or after SOURCE

(FIND-MULTICOMPILED-FILES FILES SHOWINFO) [Function]

Returns a list of files in FILES that have more than one type of compiled file (e.g. LCOM and DFASL). FILES is interpretable by FILDIR. If SHOWINFO, then the value contains a list for each file of the form

(rootname loaded-version . CREATED-AS information for each compile-type)

Otherwise just the rootname of the source is returns.

(CREATED-AS FILE) [Function]

If FILE is a Lisp source or compiled file, returns a record of its original filename and filecreated dates, and for compiled files, also the original compiled-on name and date. The return for a source file is a pair

(sfullname sfilecreateddate)

The return for a compiled file is a quadruple

(cfullname cfilecreated sfullname sfilecreateddate)

where sfullname and sourcefilecreated are extracted from the file's compiled-on information. The return is (fullname NIL) for a non-Lisp file.

(EOLTYPE FILE SHOWCONTEXT) [Function]

Returns the EOLTYPE of FILE (CR, LF, CRLF) if the type is unmistakable: contains at least one instance of one type and no instances of any others. Returns NIL if there is evidence of inconsistent types. If SHOWCONTEXT is an integer, it is the number of bytes for EOLTYPE to display before and after an instance of an inconsistent type. At each instance, the user is asked whether to continue scanning for other instances. SHOWCONTEXT = T is interpreted as 100.

(BINCOMP FILE1 FILE2 EOLDIFFOK) [Function]

Returns T if FILE1 and FILE2 are byte-identical. If EOLDIFFOK and FILE1 and FILE2 differ only in their eol conventions, the value is a list of the form (EOL1 EOL2), e.g. (CR CRLF). Otherwise the value is NIL.

