Python „C“ Utilities

Version 0.13.dev1

Franz Glasner

dev-2026-09-24

Last updated on 2026-09-24 (rev dev-7b74bd4732eb)

Contents

1 Introduction **1**
1.1 Licenses for Incorporated Software . . . 2

2 Utilities **3**
2.1 py-dos2unix . . . 3
2.2 py-genpwd . . . 5
2.3 py-shasum . . . 7
2.4 py-treesum . . . 9
2.5 py-treesum generate . . . 11
2.6 py-treesum info . . . 19

3 File Formats **21**
3.1 treesum . . . 21

4 Miscellaneous Information **37**
4.1 “treesum” Filetype Indicators . . . 37
4.2 “treesum” Filter Patterns . . . 41

5 Implementation Notes **45**
5.1 Important Observations and Facts . . . 45
5.2 Testing . . . 47

6 Changes **51**
6.1 Pre-1.0 Series . . . 51
6.2 Breaking Changes . . . 54

7 Licenses **57**

Index **59**

CHAPTER 1

Introduction

Pure Python implementations of some typical (Unix) utilities and also some additional utilities.

Metadata

Author
Franz Glasner

Version
0.13.dev1

Date
dev-2026-09-24

Copyright
© 2020–2026 Franz Glasner.

License
BSD 3-Clause “New” or “Revised” License. See BSD-3-Clause.txt for details. If you cannot find BSD-3-Clause.txt see <https://opensource.org/licenses/BSD-3-Clause>.

Parts of the code have the following licenses:

- MIT License

ID
@(#) $HGid$

Currently implemented:

- *dos2unix* as **py-dos2unix**
- *shasum* as **py-shasum**

Additional utilities:

- **py-treesum** to compute “fingerprints” or “message digests” of directory trees.
- **py-genpwd** to generate passwords of a given length and from within different character repertoires.

1.1 Licenses for Incorporated Software

crcmod/crcmod2
: MIT License

  vendored in `cutils.crcmod`

CHAPTER 2

Utilities

These are the utilities that are provided as part of the package:

2.1 py-dos2unix

Synopsis

py-dos2unix [**--help** | **-h**] [**--version** | **-V**] [**--keepdate** | **-k**] [**--oldfile** | **-o**] [**--newfile** | **-n**] [**--quiet** | **-q**] *FILE* ...

Description

Convert DOS line endings to POSIX (Unix) line endings.

Options

`--help`, **`-h`**
: Show a help message to *stdout* and exit.

`--version`, **`-V`**
: Show the program's version number to *stdout* and exit.

`--keepdate`, **`-k`**
: Keep the date stamp of the output file same as the input file.

--oldfile, **-o**

Old file mode: convert the given file and write output to it.

Wildcard names may be used safely.

The program defaults to run in this mode.

--newfile, **-n**

New file mode: convert the input file and write output to a different output file. File names must be given in pairs and wildcard names should *not* be used or you *will* loose your files.

--quiet, **-q**

Quiet mode: suppress all warnings and messages.

2.2 py-genpwd

Synopsis

py-genpwd [**--help** | **-h**] [**--version** | **-v**] [**--algorithm** | **-a** *ALGORITHM*] [**--repertoire** *REPERTOIRE*] [**-E**] [**--group** | **-G**] [**--group-sep** *GROUP-SEP*] [**--group-size** *GROUP-SIZE*] *OUTPUT-LENGTH*

Description

Generate random password from different character repertoires.

Normally *OUTPUT-LENGTH* characters from a selected character repertoire will be generated randomly.

Options

--help, **-h**
: Print a help message to *stdout* and exit.

--version, **-v**
: Print the program's version number to *stdout* and exit.

--algorithm=*ALGORITHM*, **-a** *ALGORITHM*
: Select an algorithm and (implicitly) a character repertoire.

  All algorithms that start with `bin-` just encode the output of `os.urandom()`[1] with the selected encoder. All algorithms that start with `safe-` do not contain visually similar characters (currently `0O` or `Il1`). All algorithms that end with `-2` are variants with a bias to punctuation characters.

  ALGORITHM may be one of `web`, `safe-web`, `safe-web-2`, `uri`, `safe-uri`, `safe-uri-2`, `ascii`, `safe-ascii`, `alnum`, `safe-alnum`, `bin-base64`, `bin-url-safe-base64`, `bin-base32`, `bin-hex`, `bin-ascii85`.

  Default is `safe-ascii`.

  This option is incompatible with option *--repertoire*.

-E
: For some repertoires make *OUTPUT-LENGTH* the number of bytes that is to be read from random sources instead of output bytes.

--group, **-G**
: Group the results. If no *--group-sep* or *--group-size* is given use their respective defaults.

[1] https://docs.python.org/3/library/os.html#os.urandom

--group-sep=*GROUP-SEP*

Group the result using *GROUP-SEP* as separator. Option *--group* is implied.

Default when grouping is enabled is the the space character.

--group-size=*GROUP-SIZE*

Group the result using a group size of *GROUP-SIZE* characters. Option *--group* is implied.

Default when grouping is enabled is 6.

--repertoire=*REPERTOIRE*, **-r** *REPERTOIRE*

Select from given character repertoire *REPERTOIRE*.

The repertoire must be characters from the ISO-8859-15 character set. An empty *REPERTOIRE* implicitly selects the default algorithm.

This option is incompatible with option *--algorithm*..

2.3 py-shasum

Synopsis

py-shasum [**--help** | **-h**] [**--version** | **-v**] [**--algorithm** | **-a** *ALGORITHM*] [**--base64**] [**--binary** | **-b**] [**--bsd** | **-B**] [**--check** | **-c**] [**--checklist** | **-C** *CHECKLIST*] [**--checklist-allow-distinfo**] [**--follow-symlinks**] [**--mmap**] [**--no-mmap**] [**--recurse**] [**--reverse** | **-r**] [**--tag**] [**--text** | **-t**] [*FILE* ...]

Description

Print and check SHA checksums.

With no *FILE* of when *FILE* is `-`, read from *stdin*.

Options

--help, **-h**
: Print a help message to *stdout* and exit.

--version, **-v**
: Print the program's version to *stdout* and exit.

--algorithm=*ALGORITHM*, **-a** *ALGORITHM*
: 1 (aka `sha1`), 2 (alias for `sha256`), 224, 256 (aka `sha256`), 384, 512 (aka `sha512`), 3 (alias for `sha3-512`), `3-224`, `3-256`, `3-384`, `3-512`, `blake2b`, `blake2b-256`, `blake2s`, `blake2` (alias for `blake2b`), `blake2-256` (alias for `blake2b-256`), `md5`.

  The default is 1.

--base64
: Output checksums in base64 notation, not hexadecimal (OpenBSD).

--binary, **-b**
: Read files in binary mode (default).

  Other modes are not supported!

--bsd, **-B**
: Write BSD style output. This is also the default output format of **openssl dgst**.

--check, **-c**
: Read digests from *FILE*s and check them.

  If this option is specified, the *FILE* options become checklists. Each checklist should contain hash results in a supported format, which will be verified against the specified paths. Output consists of the digest used, the file name, and an `OK`, `FAILED`, or `MISSING` for the result of the comparison. This will validate any of the supported checksums.

If no file is given, *stdin* is used.

--checklist=*CHECKLIST*, **-C** *CHECKLIST*

Compare the checksum of each *FILE* against the checksums in the *CHECKLIST*. Any specified *FILE* that is not listed in the *CHECKLIST* will generate an error.

--checklist-allow-distinfo

Allow FreeBSD *distinfo* formatted checklists: ignore SIZE and TIMESTAMP lines.

--follow-symlinks

Also follow symlinks that resolve to directories. Only effective if *--recurse* is activated.

--mmap

Use mmap[2] if available on the system.

The default depends on the size of the file that is handled.

--no-mmap

Don’t use mmap[3].

--recurse

Recurse into sub-directories while interpreting every *FILE* as a directory.

--reverse, **-r**

Explicitely select normal coreutils style output.

This is to be option compatible with BSD style commands and **openssl dgst -r**.

--tag

Alias for the *--bsd* option.

Also to be compatible with **b2sum**.

--text, **-t**

Read in text mode.

This is **not implemented**.

This option is there to be option compatible to GNU shasum.

[2] https://docs.python.org/3/library/mmap.html#module-mmap
[3] https://docs.python.org/3/library/mmap.html#module-mmap

2.4 py-treesum

Synopsis

py-treesum [**--help** | **-h**] [**--version** | **-v**] [**–debug**]

py-treesum [*GLOBAL OPTIONS*] **subcommand**

Description

Generate a “fingerprint” or “message digest” of directory trees.

Global Options

--help, **-h**
: Print a short help message to *stdout* and exit.

  This is the same as subcommand **help**.

--version, **-v**
: Print the program’s version to *stdout* and exit.

  This is the same as subcommand **version**.

--debug
: Activate debug output to *stderr*.

Subcommands

help [*SUBCOMMAND*]

: Print a short help message to *stdout* and exit.

  If no *SUBCOMMAND* is given this is the same as the global option *--help*.

  A given *SUBCOMMAND* yields the subcommand’s help and is the same as **py-treesum SUBCOMMAND --help**.

version

: Print the program’s version to *stdout* and exit.

  This is the same as the global option *--version*.

generate

: Generate “fingerprints” or “message digests” for directory trees. The output is a “treesum” file.

gen

This is an alias for the **generate** subcommand.

See *generate*.

info

Read and print “treesum” files generated by **generate**.

filetypes

Show the filetype indicators for all sorts of files.

patterns

Show the help for PATTERNs and exit.

See Also

generate, *info*, *file format*, *file types*, *filter patterns*

Bugs

- Evaluate the impact of these Python[4] changes:
  - **PEP 528**[5] (Change Windows console encoding to UTF-8, Python 3.6+)
  - **PEP 529**[6] (Change Windows filesystem encoding to UTF-8, Python 3.6+)
  - **PEP 686**[7] (Make UTF-8 mode default, Python 3.15+) and **PEP 540**[8] (Add a new UTF-8 Mode)

Todo

Test with Python 3.15!

[4] https://www.python.org/
[5] https://peps.python.org/pep-0528/
[6] https://peps.python.org/pep-0529/
[7] https://peps.python.org/pep-0686/
[8] https://peps.python.org/pep-0540/

2.5 py-treesum generate

Synopsis

py-treesum generate [**--help** | **-h**] [**--accept-treesum** | **-A** *FILTER-PATTERN*] [**--algorithm** | **-a** *ALGORITHM*] [**--append-output**] [**--base32**] [**--base64**] [**--base85**] [**--bsd**] [**--comment** *TEXT*] [**--exclude** | **-X** *FILTER-PATTERN*] [**--follow-directory-symlinks** | **-l**] [**--follow-file-symlinks**] [**--fsencoding**] [**--full-mode**] [**--generator full** | **normal** | **none**] [**--grouping-separator** *GROUPING-SEPARATOR*] [**--half** | **-H**] [**--hex**] [**--ignore-symlink-type**] [**--include** | **-I** *FILTER-PATTERN*] [**--logical** | **-L**] [**--minimal** [*TAG*]] [**--mmap**] [**--mode**] [**--mtime**] [**--no-follow-directory-symlinks**] [**--no-follow-file-symlinks**] [**--no-ignore-symlink-type**] [**--no-mmap**] [**--output** | **-o** *FILENAME*] [**--output-style** *OUTPUT-STYLE*] [**-p**] [**--physical** | **-P**] [**--print-size**] [**--reverse** | **-r**] [**--rfc1924**] [**--size-only**] [**--size-width** *SIZE-WIDTH*] [**--tag**] [**--utf8**] [**--with-notes**] *DIRECTORY* ...

Description

Generate “fingerprints” or “message digests” of every directory tree *DIRECTORY*.

Each *DIRECTORY* on the command line makes a block in the output.

The output is a “treesum” file.

> For a description of the file format see *here*.

Options

--help, **-h**
: Print the subcommand’s help message to *stdout* and exit.

--accept-treesum=*FILTER-PATTERN*, **-A** *FILTER-PATTERN*
: Accept an existing treesum file that matches *FILTER-PATTERN* for a directory fingerprint. Implicitly this also acts as *--exclude* option.

  All patterns are evaluated in given order, the first match wins.

  This option can be given more than once.

  For help on *FILTER-PATTERN* use **py-treesum patterns**.

--algorithm=*ALGORITHM*, **-a** *ALGORITHM*
: 1 (aka sha1), 2 (alias for sha256), 224, 256 (aka sha256), 384, 512 (aka sha512), 3 (alias for sha3-512), 3-224, 3-256, 3-384, 3-512, blake2b, blake2b-256, blake2s, blake2 (alias for blake2b), blake2-256 (alias for blake2b-256), md5.

  The default depends on the availability in hashlib[9]. Tested are blake2b-256, sha256 or sha1 – in this order.

--append-output
: Append to the output file instead of overwriting it.

--base32
: Output digest values in base32 notation, not hexadecimal.

--base64
: Output digest values in base64 notation, not hexadecimal.

--base85
: Output digest values in base85 notation, not hexadecimal. This encoding is compatible to the encodings used in Git[10] or Mercurial[11].

--bsd
: Alias for option *--output-style=tagged* .

--comment=*TEXT*
: Put given text *TEXT* into the output as `COMMENT` record.

  Can be given more than once.

  This option is only effective if *--minimal* is not given.

--exclude=*FILTER-PATTERN*, **-X** *FILTER-PATTERN*
: Exclude names matching the given *FILTER-PATTERN*.

  For help on *FILTER-PATTERN* use **py-treesum patterns**.

  Can be given more than once.

--follow-directory-symlinks, **-l**
: Follow symbolic links to directories when walking a directory tree.

  This option does not apply to symbolic links given on the command line.

  This option typically augments *--physical*, *--half* and *-p*.

--follow-file-symlinks
: Follow symbolic links to files when walking a directory tree.

  This is an aspect of the default behaviour.

  This option typically augments *--physical* and *--half*.

--fsencoding, **--fs-encoding**
: Use native filesystem encoding from the current locale when writing paths and texts into the output.

  This option has no effect on digest values.

  The default is to use UTF-8.

[9] https://docs.python.org/3/library/hashlib.html#module-hashlib
[10] https://git-scm.com/
[11] https://www.mercurial-scm.org/

--full-mode

Consider all mode bits as returned from *stat(2)* when computing directory message digests.

Note that mode bits on symbolic links itself are ignored. Message digests for non-directories are not affected likewise.

All the bits of `os.stat_result.st_mode`[12] are taken into account. This option most probably makes affected message digests OS dependent.

--generator=*GENERATOR*

GENERATOR may have the values `normal`, `full` or `none`.

Control the GENERATOR record in the output.

`full` prints full Python and OS/platform version information, `normal` prints just the version of **py-treesum** and whether Python 2 or Python 3 is used, and `none` suppresses the output completely.

The default is `normal`.

--grouping-separator=*GROUPING-SEPARATOR*

Use the given *GROUPING-SEPARATOR* as thousands separator when printing sizes.

The default grouping separator depends on the output style given in *--output-style*: for tagged output it is the single quote `'`, for tabbed output it is the dot `.`.

Use an empty *GROUPING-SEPARATOR* value to disable grouping.

Note that the parser recognizes just a fixed subset of values. Currently these are `.`, `,`, `'`, `_` and the space `␣`.

--half, **-H**

Follow symbolic links given on the command line but do not follow symlinks of any type while traversing the directory tree.

Overwrites any other symlink related options (*--physical*, *--logical*, *-p*, *--follow-directory-symlinks*, *--follow-file-symlinks*, et al.).

--hex

Output digest values in hex notation.

This is the default.

--ignore-symlink-type

When not folling symlinks to files and directores, also ignore the filetype of the target.

By default the target's filetype is accounted for also. So, while not following symbolic links when traversing they must be followed temporarily to determine the type of the target. This may introduce spurious inconsistencies if e.g. absolute symlinks do not exist when the digests are created and do exist when they are checked. This flag prohibits this inconsistent behaviour.

[12] https://docs.python.org/3/library/os.html#os.stat_result.st_mode

The flag is effective only when *--no-follow-directory-symlinks* and *--no-follow-file-symlinks* are both in effect, or when not following symbolic links on the command line.

This is the default.

--include=*FILTER-PATTERN*, **-I** *FILTER-PATTERN*

Include names matching the given *FILTER-PATTERN*.

All patterns are evaluated in given order, the first match wins.

This option can be given more than once.

For help on *FILTER-PATTERN* use **py-treesum patterns**.

--logical, **-L**

Follow symbolic links everywhere: on command line arguments and—while walking—directory and file symbolic links.

Overwrites any other symlink related options (*--physical*, *--half*, *-p*, *--no-follow-directory-symlinks*, *--no-follow-file-symlinks*, et al.).

--minimal[=*TAG*]

Control the content of the R00T record in the output.

By default the R00T record contains the value of the command line argument that is handled currently; additionally start and end timestamp records are generated.

With this option the R00T record will contain given *TAG* as text and no timestamp records are generated. Also any *--comment* options are ignored. A missing *TAG* counts as empty string value.

--mmap

Use `mmap`[13] if available on the system.

The default depends on the size of the file that is handled.

--mode

Consider the permission bits of *stat(2)* using `S_IMODE` (i.e. all bits without the file-type bits) when computing directory digests.

Note that mode bits on symbolic links itself are not considered.

Just some bits of `os.stat_result.st_mode`[14] are taken into account. This option may make affected message digests OS dependent.

--mtime

Consider the *mtime* of files (non-directories) when generating digests for directories. Directory entries are not considered.

Digests of files are not affected.

[13] https://docs.python.org/3/library/mmap.html#module-mmap
[14] https://docs.python.org/3/library/os.html#os.stat_result.st_mode

--no-follow-directory-symlinks

Don’t follow symbolic links to directories when walking a directory tree.

This is an aspect of the default behaviour.

This option does not apply to symbolic links given on the command line.

Also typically augments *--logical*.

--no-follow-file-symlinks

Don’t follow symbolic links to files when walking a directory tree.

--no-ignore-symlink-type

Disable the default behaviour of *--ignore-symlink-type* for phycical traversal modes.

--no-mmap

Don’t use `mmap`[15].

--output=*FILENAME*, **-o** *FILENAME*

Put the output into given file *FILENAME*.

If not given or if *FILENAME* is `-` then *stdout* is used.

--output-style=*OUTPUT-STYLE*

Select the output style or flavour:

`tagged` or `tag` select a BSD style tagged format; `tabular` or `tab` select a more GNU style (or Linux coreutils) tabular format.

Default is `tagged`.

-p

Do not follow any symbolic links to directories, whether they are given on the command line or when walking the directory tree, but follow symbolic links to files.

Overwrites any other symlink related options (*--logical*, *--half*, *--physical*, *--follow-directory-symlinks*, *--no-follow-file-symlinks*, et al.).

This is the default.

--physical, **-P**

Do not follow any symbolic links whether they are given on the command line or when walking the directory tree.

Overwrites any other symlink related options (*--logical*, *--half*, *-p*, *--follow-directory-symlinks*, *--follow-file-symlinks*, et al.).

--print-size

Print the size of a file or the accumulated sizes of directory content into the output also. The size is not considered when computing message digests.

For symbolic links the size is not printed also.

[15] https://docs.python.org/3/library/mmap.html#module-mmap

--reverse, **-r**

Alias for option *--output-style=tabular* .

--rfc1924

Output digest values in pure **RFC 1924**[16] notation.

While invented for compact representations of IPv6 addresses this encoding generates the shortest encodings of digests.

--size-only

Print only the size of files and for each directory its accumulated directory size.

--size-width=*SIZE-WIDTH*

Some output styles print the a filesize right-aligned in a column (e.g. `tabular`). *SIZE-WIDTH* is the (minimum) width to be used. The width includes grouping separators.

Use 0 if no alignment should be done.

Ignored for `tagged` output.

Default is 15.

--tag

Alias for option *--output-style=tagged* or *--bsd* .

--utf8, **--utf-8**

Use the UTF-8 encoding when writing paths and texts into the output.

Make the directory *TAINTED* if a path cannot representated in Unicode cleanly.

This is the default.

This option has no effect on digest values.

--with-notes

Also put some `NOTE` records with some extra data into the output.

Currently this is the information about excluded files.

--without-warnings

Deactivate the output of `WARNING` records. By default warnings will be generated.

Note

For symlink related options:

- These options overwrite all aspects of earlier options (or defaults): *--half*, *--logical*, *--physical*, *-p* .
- These options overwrite just some aspect of earlier options (or defaults): *--follow-directory-symlinks*, *--follow-file-symlinks*, *--no-follow-directory-symlinks*, *--no-follow-file-symlinks* .

[16] https://datatracker.ietf.org/doc/html/rfc1924.html

See Also

py-treesum, *file format filter patterns*

2.6 py-treesum info

Synopsis

py-treesum info [**–help** | **-h**] [**–last**] *TREESUM-DIGEST-FILE* ...

Description

Read and print a summary of generated “treesum” files.

For every digest block in given *TREESUM-DIGEST-FILE* print a summary with the overall block message digest, an overview over errors and other important flags and settings that have been in effect when the file was generated.

Options

--help, **-h**
: Print the subcommand’s help message to *stdout* and exit.

--last
: Print only data about the last digest block of every given file.

Examples

The file `tests/data/_data.treesum` yields this output:

```
$ python treesum.py info tests/data/_data.treesum

Tag: tests/data
FS-Encoding: UTF-8
Generator: treesum 0.12 (PY3)
Flags: denc-hex, follow-symlinks-file, no-follow-symlinks-commandline, no-
↪follow-symlinks-directory, no-ignore-symlink-type, print-size, utf8-
↪encoding
Comment: Ordinäres Treesum mit Digests und Sizes
FNMatch: exclude: glob:*.treesum
FNMatch: exclude: glob:*.size
Algorithm: SHA256
Digest: 105de9e865089039c94d3ebb7db1b9ab9ea5a55012d66ad9fc83c81dc348a1f8
Size: 639
*****TAINTED*****
Errors: directory is TAINTED
        errno 2: No such file or directory
```

The file `tests/data/_data.size` with no digests and only sizes gives this output:

```
$ python treesum.py info tests/data/_data.size

Tag: tests/data
FS-Encoding: UTF-8
Generator: treesum 0.12 (PY3)
Flags: denc-hex, follow-symlinks-file, fs-encoding, no-follow-symlinks-
↪commandline, no-follow-symlinks-directory, no-ignore-symlink-type, size-
↪only
Comment: Nur die Größen
FNMatch: exclude: glob:*.treesum
FNMatch: exclude: glob:*.size
Size: 639
*****TAINTED*****
Errors: directory is TAINTED
        errno 2: No such file or directory
```

See Also

py-treesum, *file format*

CHAPTER 3

File Formats

3.1 treesum

Description

Overall Format

The overall file format classifies as follows:

- It is a text file.
- Lines are terminated by a single U+000A character, a single U+000D character, or the two-character combination carriage return and line feed (U+000D U+000A). The last line shall be terminated likewise.

  Different newline markers can be mixed freely.
- Empty lines, lines consisting only of the whitespace characters space (U+0020) and horizontal tabulators (U+0009) are completely ignored (except for counting lines).
- Likewise ignored are (user) comment lines, lines where the first non-whitespace character is either a the number sign `#` (U+0023) or a semicolon `;` (U+003B).

  But note the difference to `COMMENT` (see *here*).
- All other lines are *records*.

  Records come in two flavours: “tagged” (BSD style) and “tabular” (GNU style). For a comparison see *this section*.
- Records are grouped into *blocks* semantically.

  A single treesum data file may contain several blocks.

Blocks

A “block” is a group of records.

Every *DIRECTORY* on the command line becomes a “block” in the output. So a “block” describes what is the essence of **`py-treesum`**: a message digest of a directory tree.

Every block starts with a `VERSION` record and regularly ends with a `CRC32` record. If the `CRC32` record is missing a block ends at EOF or at the next `VERSION` record in the data—which ends the current block and starts a new block.

The primary encoding of a block is either UTF-8 or the encoding given in the block's `FSENCODING` record. This is documented in the `FLAGS` record. Note that the primary encoding of the content may differ from block to block.

All records of a block must share a consistent single flavour “tagged” or “tabular”. Different blocks may have different record flavours.

Records in Lines: “tagged” vs “tabular”

> **Note**
>
> Horizontal tabular characters (U+0009) are printed here as “␉” glyphs.

In the “tabular” format each record consists of two or more TAB-separated columns. The record's name or type may be derived from the number of colums or other context data.

In the “tagged” format each record (aka line) begins with the record's name or type, which is unique.

Record *NAME*	“tagged”	“tabular”
VERSION, CRC32, FSENCODING, GENERATOR, FLAGS, START-TIMESTAMP, END-TIMESTAMP, COMMENT, FNMATCH	*NAME* = *VALUE*	*NAME* ⭾ *VALUE*
ROOT	ROOT(*PATH*)	ROOT ⭾ *PATH*
ACCEPT-TREESUM	ACCEPT-TREESUM(*PATH*) = *FILE*	ACCEPT-TREESUM:*FILE* ⭾ *{PATH}* ACCEPT-TREESUM:*FILE* ⭾ ⭾ *{PATH}*
ERROR	ERROR(*PATH*) = *MESSAGE*	ERR:*MESSAGE* ⭾ *PATH* ERR:*MESSAGE* ⭾ ///// ⭾ *PATH*
WARNING	WARNING(*PATH*) = *MESSAGE*	WARN:*MESSAGE* ⭾ *PATH* WARN:*MESSAGE* ⭾ ///// ⭾ *PATH*
NOTE	NOTE(*PATH*) = *MESSAGE*	NOTE:*MESSAGE* ⭾ *PATH* NOTE:*MESSAGE* ⭾ ///// ⭾ *PATH*
SIZE	SIZE(*PATH*) = *SZ* SIZE(*PATH*)	*SZ* ⭾ *PATH*
DIGEST	*DIGEST*(*PATH*) = *VALUE* *DIGEST*(*PATH*) = *VALUE*,*SZ* *DIGEST*(*PATH*)	*DIGEST*:*VALUE* ⭾ *PATH* *DIGEST*:*VALUE* ⭾ *SZ* ⭾ *PATH*

Some additional notes:

- In “tabular” flavour all *SZ* values may be numbers, or—in error cases—a sequence of slash characters / (e.g. ////////).
- In “tagged” style all *SZ* values also may be the empty string.

Records

`VERSION`

Required to start a block and to declare its format version.

Currently the version value is 1.

The current block regularly ends with a `CRC32` record or—if this is missing—at EOF or the next `VERSION` record in the data.

`CRC32`

The CRC checksum regularly ends the current block.

It is the hex-encoded CRC-32[17] (as used in Ethernet, ZIP, PNG et al.) of all the lines of the block as seen in the input, using the normalized newline character Unicode code point U+000A on all operating systems.

(User) comment lines (first non-whitespace is `#` or `;`), lines consisting of whitespace characters (tabs and spaces) only and empty lines do not contribute to the CRC. These lines are ignored with the exception of contributing to line numbering.

`FSENCODING`

The result of `sys.getfilesystemencoding()`[18] that is in effect when generating treesum digests (process environment)

`GENERATOR`

The existence of this record and its value depends on the option *`--generator`*:

`full` yields full Python and OS/platform version information.

`normal` gives just the version of py-treesum and whether Python 2 or Python 3 is used.

With `none` the record will not be found in the output at all.

`FLAGS`

A comma-separated list of flag values (strings).

They reflect important settings that were in effect when the current block has been generated.

[17] https://en.wikipedia.org/wiki/CRC-32
[18] https://docs.python.org/3/library/sys.html#sys.getfilesystemencoding

Possible Flag Value	Always	Description
`denc-`*`ENCODING`*	√	Encoding of message digests. Possible values of *ENCODING* are `hex`, `base32`, `base64`, `base85` and `rfc1924`.
`fs-encoding`, `utf8-encoding`	√	Denote whether either `sys.getfilesystemencoding()`[19] (as recorded in the `FSENCODING` record) or `utf-8` are the primary encoding of all the records of the current block. This affects primarily path names, but also all other “text” values like error messages, filter expressions and comment records. Message digest values are *not* affected.
`follow-symlinks-commandline`, `no-follow-symlinks-commandline`	√	Symlink handling for command line arguments
`follow-symlinks-directory`, `no-follow-symlinks-directory`	√	Symlink handling for directories
`follow-symlinks-file`, `no-follow-symlinks-file`	√	Symlink handling for non-directories
`ignore-symlink-type`, `no-ignore-symlink-type`	√	Reflect whether *`--ignore-symlink-type`* or *`--no-ignore-symlink-type`* was in effect when the treesum block data was generated.
`print-size`		*`--print-size`* was also given and—additionally to message digests—file sizes and accumulated file sizes for directory trees are in the block data; the values of message digests are not affected in any way.
`size-only`		*`--size-only`* was given: no digests but only file sizes and accumulated file sizes for directory trees are in the block’s data.
`with-metadata-fullmode`		*`--full-mode`* was in effect: all all mode bits as returned from *`stat(2)`* were used in digest computations for directory message digests.
`with-metadata-mode`		*`--mode`* was in effect: the permission bits of *`stat(2)`* using `S_IMODE` (i.e. all bits without the filetype bits) contributed to directory message digests.
`with-metadata-mtime`		*`--mtime`* was in effect: the *mtime* as ISO string formatted timestamp (truncated to seconds) was used in message digest computations also.
`with-notes`		*`--with-notes`* was in effect
`without-warnings`		*`--without-warnings`* was in effect

A “√” character in the “Always” column denotes a flag where one of its possible

[19] https://docs.python.org/3/library/sys.html#sys.getfilesystemencoding

values is always there. Otherwise a “missing” flag denotes a default value.

`START-TIMESTAMP`, `END-TIMESTAMP`

Contain the start and end times of **`py-treesum generate`** in Unix epoch time and as ISO formatted string (truncated to seconds).

These records are generated only if *`--minimal`* is not given.

`COMMENT`

Contains the value of a *`--comment`* option.

Its existence depends on *`--minimal`*.

`FNMATCH`

Remembers the option values for options *`--include`*, *`--exclude`* and *`--accept-treesum`*.

`ROOT`

By default contains the root of the block’s directory tree as given on the command line—but somewhat normalized.

If *`--minimal=VALUE`* is used then *VALUE* is put into the `ROOT` record.

`SIZE`

If *`--print-size`* is used it replaces *`DIGEST`* records. See also *there*.

`ERROR`, `WARNING`, `NOTE`

These records contain non-fatal errors, warnings or other notes for a given *PATH* as text messages. Their individual message formats are somewhat unspecific.

`ACCEPT-TREESUM`

Note that a message digest has been accepted for a given *PATH* from a within file *FILE*.

Precedes the accepted message digest or size record.

`DIGEST`

The last digest record is the aggregated message digest of the current block. Its *PATH* values and file types differ from the other file types given in other records.

For more details see *file types*.

Note that this is true also for associated records (e.g. `ERROR`, `WARNING`, `NOTE`, `ACCEPT-TREESUM`).

Examples

py-treesum generate -P -a md5 --base85 --utf8 --print-size --with-notes -X $'re:\\.(treesum|size)\\Z' yields:

```
VERSION = 1
FSENCODING = UTF-8
GENERATOR = treesum 0.13.dev1 (PY3)
FLAGS = denc-base85,ignore-symlink-type,no-follow-symlinks-commandline,no-
↪follow-symlinks-directory,no-follow-symlinks-file,print-size,utf8-
↪encoding,with-notes
START-TIMESTAMP = 1789643152;2026-09-17T11:05:52+00:00
COMMENT = These are comment records ...
COMMENT = ... with german umlauts: ÄÖÜäöüß
FNMATCH = exclude: re:\.(treesum|size)\Z
ROOT (tests/data)
NOTE (_data-tabular.size) = SKIPPED
NOTE (_data-tabular.treesum) = SKIPPED
NOTE (_data.size) = SKIPPED
NOTE (_data.treesum) = SKIPPED
MD5 (cp1252/ED K?hlen Unterbau integrierbar.txt) = rdj)Q97HI{^?bhm%WeIQ,
↪286
MD5 (cp1252/) = pjtI_`Yx-{HTq^}yqTnq,286
MD5 (empty) = )E$i3j{vd+>3|8C?D#@{,0
MD5 (tree-1/file-1-1-s.txt/./@@) = xA6W44p|6H{6tP9JprW>,
MD5 (tree-1/file-1-1.txt) = @AbSM`0Skn=0N;IaP{iV,3
MD5 (tree-1/file-1-2.txt) = <QM)?xn^5^_Up8I<;!xk,5
MD5 (tree-1/file-1-nn-s.txt/./@@) = F;c0D<LA(a2Mj(WrL%Wo,
MD5 (tree-1/) = `NK5R_pX){M<jO6yFOxM,8
MD5 (tree-1-s/./@@) = -rMfx(}?S@lFsC9IN}!m,
MD5 (tree-2/empty-1) = )E$i3j{vd+>3|8C?D#@{,0
MD5 (tree-2/tree-1-s/./@@) = sl4wU;GXSMxPghs&)C&o,
MD5 (tree-2/tree-nn-s/./@@) = mFx)c&pW$k?~hS^q4hE$,
MD5 (tree-2/) = y+*qfv6XmbQvYh$fIrLK,0
MD5 (tree-3/file-3-1.txt) = fhu7XpG26sQ_f0!I@JzX,7
MD5 (tree-3/file-3-2.txt) = 4Ty2f_BHQFdx)=lBhCv;,10
MD5 (tree-3/tree-3-1/file-3-1-1.txt) = @AbSM`0Skn=0N;IaP{iV,3
MD5 (tree-3/tree-3-1/file-3-1-2-s.txt/./@@) = jxXVFC38i7tb=hEE69m`,
MD5 (tree-3/tree-3-1/file-3-1-2.txt) = dj!jZf05z5j#3HG%d1Zm,9
MD5 (tree-3/tree-3-1/) = i7uW7{WVGHq0a8D<1)G6,12
MD5 (tree-3/tree-3-1-s/./@@) = ;~R1N3IuJ%8mBVV5L%Jq,
MD5 (tree-3/) = CM3DuI89j$0)?_!K%Q^P,29
MD5 (tree-3-2/file-3-2-1.txt) = 9{WI}+N+I|usA%X%%pia,9
MD5 (tree-3-2/file-3-2-2.txt) = 8V#2oGG88-Mqa0$?)0x<,9
MD5 (tree-3-2/) = z|rebkV4<IP>7{KFe4o+,18
MD5 (tree-nn-s/./@@) = >2z{C+9T3isrfNL)Xwny,
MD5 (utf-8/ED Kühlen Unterbau integrierbar.txt) = rdj)Q97HI{^?bhm%WeIQ,286
MD5 (utf-8/) = $Wrg`kGIKcrARWc`}^&i,286
MD5 () = cM@)n*}De!OLW`ZbhhMv,627
END-TIMESTAMP = 1789643152;2026-09-17T11:05:52+00:00
CRC32 = E265C939
```

The same directory tree in the tabular output format using **py-treesum generate -P -a md5 --base85 --utf8 --print-size --with-notes -X $'re:\\.(treesum|size)\\Z' --reverse --size-width=4** yields:

```
VERSION        1
FSENCODING        UTF-8
GENERATOR        treesum 0.13.dev1 (PY3)
FLAGS        denc-base85,ignore-symlink-type,no-follow-symlinks-
↪commandline,no-follow-symlinks-directory,no-follow-symlinks-file,print-
↪size,utf8-encoding,with-notes
START-TIMESTAMP        1789643152;2026-09-17T11:05:52+00:00
COMMENT        These are comment records ...
COMMENT        ... with german umlauts: ÄÖÜäöüß
FNMATCH        exclude: re:\.(treesum|size)\Z
ROOT        tests/data
NOTE:SKIPPED                    ////        _data-tabular.size
NOTE:SKIPPED                    ////        _data-tabular.treesum
NOTE:SKIPPED                    ////        _data.size
NOTE:SKIPPED                    ////        _data.treesum
MD5:rdj)Q97HI{^?bhm%WeIQ         286        cp1252/ED K?hlen Unterbau␣
↪integrierbar.txt
MD5:pjtI_`Yx-{HTq^}yqTnq         286        cp1252/
MD5:)E$i3j{vd+>3|8C?D#@{           0        empty
MD5:xA6W44p|6H{6tP9JprW>                    tree-1/file-1-1-s.txt/./@@
MD5:@AbSM`0Skn=0N;IaP{iV           3        tree-1/file-1-1.txt
MD5:<QM)?xn^5^_Up8I<;!xk           5        tree-1/file-1-2.txt
MD5:F;c0D<LA(a2Mj(WrL%Wo                    tree-1/file-1-nn-s.txt/./@@
MD5:`NK5R_pX){M<jO6yFOxM           8        tree-1/
MD5:-rMfx(}?S@lFsC9IN}!m                    tree-1-s/./@@
MD5:)E$i3j{vd+>3|8C?D#@{           0        tree-2/empty-1
MD5:sl4wU;GXSMxPghs&)C&o                    tree-2/tree-1-s/./@@
MD5:mFx)c&pW$k?~hS^q4hE$                    tree-2/tree-nn-s/./@@
MD5:y+*qfv6XmbQvYh$fIrLK           0        tree-2/
MD5:fhu7XpG26sQ_f0!I@JzX           7        tree-3/file-3-1.txt
MD5:4Ty2f_BHQFdx)=lBhCv;          10        tree-3/file-3-2.txt
MD5:@AbSM`0Skn=0N;IaP{iV           3        tree-3/tree-3-1/file-3-1-1.txt
MD5:jxXVFC38i7tb=hEE69m`                    tree-3/tree-3-1/file-3-1-2-s.
↪txt/./@@
MD5:dj!jZf05z5j#3HG%d1Zm           9        tree-3/tree-3-1/file-3-1-2.txt
MD5:i7uW7{WVGHq0a8D<1)G6          12        tree-3/tree-3-1/
MD5:;~R1N3IuJ%8mBVV5L%Jq                    tree-3/tree-3-1-s/./@@
MD5:CM3DuI89j$0)?_!K%Q^P          29        tree-3/
MD5:9{WI}+N+I|usA%X%%pia           9        tree-3-2/file-3-2-1.txt
MD5:8V#2oGG88-Mqa0$?)0x<           9        tree-3-2/file-3-2-2.txt
MD5:z|rebkV4<IP>7{KFe4o+          18        tree-3-2/
MD5:>2z{C+9T3isrfNL)Xwny                    tree-nn-s/./@@
MD5:rdj)Q97HI{^?bhm%WeIQ         286        utf-8/ED Kühlen Unterbau␣
↪integrierbar.txt
MD5:$Wrg`kGIKcrARWc`}^&i         286        utf-8/
MD5:cM@)n*}De!OLW`ZbhhMv         627        .
```

(continues on next page)

(continued from previous page)

```
END-TIMESTAMP        1789643152;2026-09-17T11:05:52+00:00
CRC32        9FE07B69
```

Logical traversal with resolving all symbolic links should give errors.

py-treesum generate -L -a md5 --base85 --utf8 --print-size --with-notes -X $'re:\\.(treesum|size)\\Z' :

```
VERSION = 1
FSENCODING = UTF-8
GENERATOR = treesum 0.13.dev1 (PY3)
FLAGS = denc-base85,follow-symlinks-commandline,follow-symlinks-directory,
↪follow-symlinks-file,ignore-symlink-type,print-size,utf8-encoding,with-
↪notes
START-TIMESTAMP = 1789643152;2026-09-17T11:05:52+00:00
COMMENT = These are comment records ...
COMMENT = ... with german umlauts: ÄÖÜäöüß
FNMATCH = exclude: re:\.(treesum|size)\Z
ROOT (tests/data)
NOTE (_data-tabular.size) = SKIPPED
NOTE (_data-tabular.treesum) = SKIPPED
NOTE (_data.size) = SKIPPED
NOTE (_data.treesum) = SKIPPED
MD5 (cp1252/ED K?hlen Unterbau integrierbar.txt) = rdj)Q97HI{^?bhm%WeIQ,
↪286
MD5 (cp1252/) = pjtI_`Yx-{HTq^}yqTnq,286
MD5 (empty) = )E$i3j{vd+>3|8C?D#@{,0
MD5 (tree-1/file-1-1-s.txt) = @AbSM`0Skn=0N;IaP{iV,3
MD5 (tree-1/file-1-1.txt) = @AbSM`0Skn=0N;IaP{iV,3
MD5 (tree-1/file-1-2.txt) = <QM)?xn^5^_Up8I<;!xk,5
ERROR (tree-1/file-1-nn-s.txt) = errno 2: No such file or directory
MD5 (tree-1/file-1-nn-s.txt)
ERROR (tree-1/) = directory is TAINTED
MD5 (tree-1/) = =#gafTrUE9_}PX%MQn$a,11
MD5 (tree-1-s/file-1-1-s.txt) = @AbSM`0Skn=0N;IaP{iV,3
MD5 (tree-1-s/file-1-1.txt) = @AbSM`0Skn=0N;IaP{iV,3
MD5 (tree-1-s/file-1-2.txt) = <QM)?xn^5^_Up8I<;!xk,5
ERROR (tree-1-s/file-1-nn-s.txt) = errno 2: No such file or directory
MD5 (tree-1-s/file-1-nn-s.txt)
ERROR (tree-1-s/) = directory is TAINTED
MD5 (tree-1-s/) = =#gafTrUE9_}PX%MQn$a,11
MD5 (tree-2/empty-1) = )E$i3j{vd+>3|8C?D#@{,0
MD5 (tree-2/tree-1-s/file-1-1-s.txt) = @AbSM`0Skn=0N;IaP{iV,3
MD5 (tree-2/tree-1-s/file-1-1.txt) = @AbSM`0Skn=0N;IaP{iV,3
MD5 (tree-2/tree-1-s/file-1-2.txt) = <QM)?xn^5^_Up8I<;!xk,5
ERROR (tree-2/tree-1-s/file-1-nn-s.txt) = errno 2: No such file or␣
↪directory
MD5 (tree-2/tree-1-s/file-1-nn-s.txt)
ERROR (tree-2/tree-1-s/) = directory is TAINTED
```

(continues on next page)

(continued from previous page)

```
MD5 (tree-2/tree-1-s/) = =#gafTrUE9_}PX%MQn$a,11
ERROR (tree-2/tree-nn-s) = errno 2: No such file or directory
MD5 (tree-2/tree-nn-s)
ERROR (tree-2/) = directory is TAINTED
MD5 (tree-2/) = adrm!^zJ%ugryYci#Un2,11
MD5 (tree-3/file-3-1.txt) = fhu7XpG26sQ_f0!I@JzX,7
MD5 (tree-3/file-3-2.txt) = 4Ty2f_BHQFdx)=lBhCv;,10
MD5 (tree-3/tree-3-1/file-3-1-1.txt) = @AbSM`0Skn=0N;IaP{iV,3
MD5 (tree-3/tree-3-1/file-3-1-2-s.txt) = dj!jZf05z5j#3HG%d1Zm,9
MD5 (tree-3/tree-3-1/file-3-1-2.txt) = dj!jZf05z5j#3HG%d1Zm,9
MD5 (tree-3/tree-3-1/) = u^fJ=eCeXqi!AkaFtPQH,21
MD5 (tree-3/tree-3-1-s/file-3-1-1.txt) = @AbSM`0Skn=0N;IaP{iV,3
MD5 (tree-3/tree-3-1-s/file-3-1-2-s.txt) = dj!jZf05z5j#3HG%d1Zm,9
MD5 (tree-3/tree-3-1-s/file-3-1-2.txt) = dj!jZf05z5j#3HG%d1Zm,9
MD5 (tree-3/tree-3-1-s/) = u^fJ=eCeXqi!AkaFtPQH,21
MD5 (tree-3/) = 8zx|;lgJ(!6?1{{)<cr{,59
MD5 (tree-3-2/file-3-2-1.txt) = 9{WI}+N+I|usA%X%%pia,9
MD5 (tree-3-2/file-3-2-2.txt) = 8V#2oGG88-Mqa0$?)0x<,9
MD5 (tree-3-2/) = z|rebkV4<IP>7{KFe4o+,18
ERROR (tree-nn-s) = errno 2: No such file or directory
MD5 (tree-nn-s)
MD5 (utf-8/ED Kühlen Unterbau integrierbar.txt) = rdj)Q97HI{^?bhm%WeIQ,286
MD5 (utf-8/) = $Wrg`kGIKcrARWc`}^&i,286
ERROR () = directory is TAINTED
MD5 () = ZW|qu`w&7sY2+F;==OsA,682
END-TIMESTAMP = 1789643152;2026-09-17T11:05:52+00:00
CRC32 = 28EF9843
```

py-treesum generate -L -a md5 --base85 --utf8 --print-size --with-notes -X $'re:\\.(treesum|size)\\Z' --reverse --size-width=4:

```
VERSION        1
FSENCODING        UTF-8
GENERATOR        treesum 0.13.dev1 (PY3)
FLAGS        denc-base85,follow-symlinks-commandline,follow-symlinks-
↪directory,follow-symlinks-file,ignore-symlink-type,print-size,utf8-
↪encoding,with-notes
START-TIMESTAMP        1789643152;2026-09-17T11:05:52+00:00
COMMENT        These are comment records ...
COMMENT        ... with german umlauts: ÄÖÜäöüß
FNMATCH        exclude: re:\.(treesum|size)\Z
ROOT        tests/data
NOTE:SKIPPED                    ////        _data-tabular.size
NOTE:SKIPPED                    ////        _data-tabular.treesum
NOTE:SKIPPED                    ////        _data.size
NOTE:SKIPPED                    ////        _data.treesum
MD5:rdj)Q97HI{^?bhm%WeIQ         286        cp1252/ED K?hlen Unterbau␣
↪integrierbar.txt
MD5:pjtI_`Yx-{HTq^}yqTnq         286        cp1252/
```

(continues on next page)

(continued from previous page)

```
MD5:)E$i3j{vd+>3|8C?D#@{           0        empty
MD5:@AbSM`0Skn=0N;IaP{iV           3        tree-1/file-1-1-s.txt
MD5:@AbSM`0Skn=0N;IaP{iV           3        tree-1/file-1-1.txt
MD5:<QM)?xn^5^_Up8I<;!xk           5        tree-1/file-1-2.txt
ERR:errno 2: No such file or directory        ////        tree-1/file-1-
↪nn-s.txt
MD5:////////////////////        ////        tree-1/file-1-nn-s.txt
ERR:directory is TAINTED        ////        tree-1/
MD5:=#gafTrUE9_}PX%MQn$a          11        tree-1/
MD5:@AbSM`0Skn=0N;IaP{iV           3        tree-1-s/file-1-1-s.txt
MD5:@AbSM`0Skn=0N;IaP{iV           3        tree-1-s/file-1-1.txt
MD5:<QM)?xn^5^_Up8I<;!xk           5        tree-1-s/file-1-2.txt
ERR:errno 2: No such file or directory        ////        tree-1-s/file-1-
↪nn-s.txt
MD5:////////////////////        ////        tree-1-s/file-1-nn-s.txt
ERR:directory is TAINTED        ////        tree-1-s/
MD5:=#gafTrUE9_}PX%MQn$a          11        tree-1-s/
MD5:)E$i3j{vd+>3|8C?D#@{           0        tree-2/empty-1
MD5:@AbSM`0Skn=0N;IaP{iV           3        tree-2/tree-1-s/file-1-1-s.txt
MD5:@AbSM`0Skn=0N;IaP{iV           3        tree-2/tree-1-s/file-1-1.txt
MD5:<QM)?xn^5^_Up8I<;!xk           5        tree-2/tree-1-s/file-1-2.txt
ERR:errno 2: No such file or directory        ////        tree-2/tree-1-s/
↪file-1-nn-s.txt
MD5:////////////////////        ////        tree-2/tree-1-s/file-1-nn-s.
↪txt
ERR:directory is TAINTED        ////        tree-2/tree-1-s/
MD5:=#gafTrUE9_}PX%MQn$a          11        tree-2/tree-1-s/
ERR:errno 2: No such file or directory        ////        tree-2/tree-nn-s
MD5:////////////////////        ////        tree-2/tree-nn-s
ERR:directory is TAINTED        ////        tree-2/
MD5:adrm!^zJ%ugryYci#Un2          11        tree-2/
MD5:fhu7XpG26sQ_f0!I@JzX           7        tree-3/file-3-1.txt
MD5:4Ty2f_BHQFdx)=lBhCv;          10        tree-3/file-3-2.txt
MD5:@AbSM`0Skn=0N;IaP{iV           3        tree-3/tree-3-1/file-3-1-1.txt
MD5:dj!jZf05z5j#3HG%d1Zm           9        tree-3/tree-3-1/file-3-1-2-s.
↪txt
MD5:dj!jZf05z5j#3HG%d1Zm           9        tree-3/tree-3-1/file-3-1-2.txt
MD5:u^fJ=eCeXqi!AkaFtPQH          21        tree-3/tree-3-1/
MD5:@AbSM`0Skn=0N;IaP{iV           3        tree-3/tree-3-1-s/file-3-1-1.
↪txt
MD5:dj!jZf05z5j#3HG%d1Zm           9        tree-3/tree-3-1-s/file-3-1-2-
↪s.txt
MD5:dj!jZf05z5j#3HG%d1Zm           9        tree-3/tree-3-1-s/file-3-1-2.
↪txt
MD5:u^fJ=eCeXqi!AkaFtPQH          21        tree-3/tree-3-1-s/
MD5:8zx|;lgJ(!6?1{{)<cr{          59        tree-3/
MD5:9{WI}+N+I|usA%X%%pia           9        tree-3-2/file-3-2-1.txt
MD5:8V#2oGG88-Mqa0$?)0x<           9        tree-3-2/file-3-2-2.txt
MD5:z|rebkV4<IP>7{KFe4o+          18        tree-3-2/
```

(continues on next page)

(continued from previous page)

```
ERR:errno 2: No such file or directory        ////        tree-nn-s
MD5://////////////////////        ////        tree-nn-s
MD5:rdj)Q97HI{^?bhm%WeIQ         286        utf-8/ED Kühlen Unterbau␣
↪integrierbar.txt
MD5:$Wrg`kGIKcrARWc`}^&i         286        utf-8/
ERR:directory is TAINTED        ////        .
MD5:ZW|qu`w&7sY2+F;==OsA         682        .
END-TIMESTAMP        1789643152;2026-09-17T11:05:52+00:00
CRC32        F73EB8FA
```

Printing only the sizes gives these results ...

... for tagged output using **py-treesum generate -L -a md5 --base85 --utf8 --size-only --with-notes -X $'re:\\.(treesum|size)\\Z'** :

```
VERSION = 1
FSENCODING = UTF-8
GENERATOR = treesum 0.13.dev1 (PY3)
FLAGS = denc-base85,follow-symlinks-commandline,follow-symlinks-directory,
↪follow-symlinks-file,ignore-symlink-type,size-only,utf8-encoding,with-
↪notes
START-TIMESTAMP = 1789643152;2026-09-17T11:05:52+00:00
COMMENT = These are comment records ...
COMMENT = ... with german umlauts: ÄÖÜäöüß
FNMATCH = exclude: re:\.(treesum|size)\Z
ROOT (tests/data)
NOTE (_data-tabular.size) = SKIPPED
NOTE (_data-tabular.treesum) = SKIPPED
NOTE (_data.size) = SKIPPED
NOTE (_data.treesum) = SKIPPED
SIZE (cp1252/ED K?hlen Unterbau integrierbar.txt) = 286
SIZE (cp1252/) = 286
SIZE (empty) = 0
SIZE (tree-1/file-1-1-s.txt) = 3
SIZE (tree-1/file-1-1.txt) = 3
SIZE (tree-1/file-1-2.txt) = 5
ERROR (tree-1/file-1-nn-s.txt) = errno 2: No such file or directory
SIZE (tree-1/file-1-nn-s.txt)
ERROR (tree-1/) = directory is TAINTED
SIZE (tree-1/) = 11
SIZE (tree-1-s/file-1-1-s.txt) = 3
SIZE (tree-1-s/file-1-1.txt) = 3
SIZE (tree-1-s/file-1-2.txt) = 5
ERROR (tree-1-s/file-1-nn-s.txt) = errno 2: No such file or directory
SIZE (tree-1-s/file-1-nn-s.txt)
ERROR (tree-1-s/) = directory is TAINTED
SIZE (tree-1-s/) = 11
SIZE (tree-2/empty-1) = 0
SIZE (tree-2/tree-1-s/file-1-1-s.txt) = 3
```

(continues on next page)

(continued from previous page)

```
SIZE (tree-2/tree-1-s/file-1-1.txt) = 3
SIZE (tree-2/tree-1-s/file-1-2.txt) = 5
ERROR (tree-2/tree-1-s/file-1-nn-s.txt) = errno 2: No such file or␣
↪directory
SIZE (tree-2/tree-1-s/file-1-nn-s.txt)
ERROR (tree-2/tree-1-s/) = directory is TAINTED
SIZE (tree-2/tree-1-s/) = 11
ERROR (tree-2/tree-nn-s) = errno 2: No such file or directory
SIZE (tree-2/tree-nn-s)
ERROR (tree-2/) = directory is TAINTED
SIZE (tree-2/) = 11
SIZE (tree-3/file-3-1.txt) = 7
SIZE (tree-3/file-3-2.txt) = 10
SIZE (tree-3/tree-3-1/file-3-1-1.txt) = 3
SIZE (tree-3/tree-3-1/file-3-1-2-s.txt) = 9
SIZE (tree-3/tree-3-1/file-3-1-2.txt) = 9
SIZE (tree-3/tree-3-1/) = 21
SIZE (tree-3/tree-3-1-s/file-3-1-1.txt) = 3
SIZE (tree-3/tree-3-1-s/file-3-1-2-s.txt) = 9
SIZE (tree-3/tree-3-1-s/file-3-1-2.txt) = 9
SIZE (tree-3/tree-3-1-s/) = 21
SIZE (tree-3/) = 59
SIZE (tree-3-2/file-3-2-1.txt) = 9
SIZE (tree-3-2/file-3-2-2.txt) = 9
SIZE (tree-3-2/) = 18
ERROR (tree-nn-s) = errno 2: No such file or directory
SIZE (tree-nn-s)
SIZE (utf-8/ED Kühlen Unterbau integrierbar.txt) = 286
SIZE (utf-8/) = 286
ERROR () = directory is TAINTED
SIZE () = 682
END-TIMESTAMP = 1789643152;2026-09-17T11:05:52+00:00
CRC32 = F18A7519
```

... for tabular output using **py-treesum generate -L -a md5 --base85 --utf8 --size-only --with-notes -X $'re:\\.(treesum|size)\\Z' --reverse --size-width=4**:

```
VERSION        1
FSENCODING        UTF-8
GENERATOR        treesum 0.13.dev1 (PY3)
FLAGS        denc-base85,follow-symlinks-commandline,follow-symlinks-
↪directory,follow-symlinks-file,ignore-symlink-type,size-only,utf8-
↪encoding,with-notes
START-TIMESTAMP        1789643152;2026-09-17T11:05:52+00:00
COMMENT        These are comment records ...
COMMENT        ... with german umlauts: ÄÖÜäöüß
FNMATCH        exclude: re:\.(treesum|size)\Z
ROOT        tests/data
```

(continues on next page)

(continued from previous page)

```
NOTE:SKIPPED                    _data-tabular.size
NOTE:SKIPPED                    _data-tabular.treesum
NOTE:SKIPPED                    _data.size
NOTE:SKIPPED                    _data.treesum
 286        cp1252/ED K?hlen Unterbau integrierbar.txt
 286        cp1252/
   0        empty
   3        tree-1/file-1-1-s.txt
   3        tree-1/file-1-1.txt
   5        tree-1/file-1-2.txt
ERR:errno 2: No such file or directory        tree-1/file-1-nn-s.txt
////        tree-1/file-1-nn-s.txt
ERR:directory is TAINTED        tree-1/
  11        tree-1/
   3        tree-1-s/file-1-1-s.txt
   3        tree-1-s/file-1-1.txt
   5        tree-1-s/file-1-2.txt
ERR:errno 2: No such file or directory        tree-1-s/file-1-nn-s.txt
////        tree-1-s/file-1-nn-s.txt
ERR:directory is TAINTED        tree-1-s/
  11        tree-1-s/
   0        tree-2/empty-1
   3        tree-2/tree-1-s/file-1-1-s.txt
   3        tree-2/tree-1-s/file-1-1.txt
   5        tree-2/tree-1-s/file-1-2.txt
ERR:errno 2: No such file or directory        tree-2/tree-1-s/file-1-nn-s.
↪txt
////        tree-2/tree-1-s/file-1-nn-s.txt
ERR:directory is TAINTED        tree-2/tree-1-s/
  11        tree-2/tree-1-s/
ERR:errno 2: No such file or directory        tree-2/tree-nn-s
////        tree-2/tree-nn-s
ERR:directory is TAINTED        tree-2/
  11        tree-2/
   7        tree-3/file-3-1.txt
  10        tree-3/file-3-2.txt
   3        tree-3/tree-3-1/file-3-1-1.txt
   9        tree-3/tree-3-1/file-3-1-2-s.txt
   9        tree-3/tree-3-1/file-3-1-2.txt
  21        tree-3/tree-3-1/
   3        tree-3/tree-3-1-s/file-3-1-1.txt
   9        tree-3/tree-3-1-s/file-3-1-2-s.txt
   9        tree-3/tree-3-1-s/file-3-1-2.txt
  21        tree-3/tree-3-1-s/
  59        tree-3/
   9        tree-3-2/file-3-2-1.txt
   9        tree-3-2/file-3-2-2.txt
  18        tree-3-2/
ERR:errno 2: No such file or directory        tree-nn-s
```

(continues on next page)

(continued from previous page)

```
////        tree-nn-s
 286        utf-8/ED Kühlen Unterbau integrierbar.txt
 286        utf-8/
ERR:directory is TAINTED        .
 682        .
END-TIMESTAMP        1789643152;2026-09-17T11:05:52+00:00
CRC32         D9B25DFC
```

See Also

py-treesum, *file type indicators*

Bugs

Version 1 of the file format may change in versions of **py-treesum** before version `1.0`; backwards compatibility is not really retained before version `1.0`.

CHAPTER 4

Miscellaneous Information

4.1 “treesum” Filetype Indicators

Description

File and directory paths are printed using names similar to calling **`ls -F`** or **`ls --classify`**. The indicator strings (aka “marker” or “tags”) are appended to their names as follows.

In an aggregated directory message digest (i.e. the last digest record at the end of a block)—if the path *equals*:

`.` or an empty string

Normal aggregated message digest and/or size.

The dot `.` is used in tabular output style, the empty string in the tagged style.

`./@@`

Symlink to some filesystem object; the target type is ignored.

Used in physical mode (i.e when not following symlinks for commandline arguments) if the target type is willently ignored. Therefore the link may be broken or not.

This marker is seen only when option *`--ignore-symlink-type`* is in effect.

`./@/`

Symlink to directory.

`./@`

Symlink to other filesystem object. Also used if the link is broken and the target type cannot determined.

“Normal” paths that *end* with:

`/./@@`

Symlink to some filesystem object; the target type is ignored.

Used in physical mode (i.e when not following symlinks to files and/or directories) if the target type is willently ignored. Therefore the link may be broken or not.

This marker is only used when options *--no-follow-directory-symlinks*, *--no-follow-file-symlinks* and *--ignore-symlink-type* are in effect.

`/`

Denotes a directory.

`/./@/`

Denotes a Symlink to a directory.

`/./@`

Classifies a symlink to a regular filesystem object if the target object is not classified as directoriy or other special filesystem object otherwise.

Also used if the link is broken and the target type cannot determined.

Seen only if *--no-ignore-symlink-type* is in effect and target types should not ignored.

`/./|`, `/./@|`

FIFO and symlink to FIFO.

`/./=`, `/./@=`

Socket and symlink to socket.

`/./>`, `/./@>`

Door (Solaris) and symlink to door.

`/./%`, `/./@%`

Whiteout and symlink to whiteout (used by union filesystems) (BSD).

Other non-standard path *endings* are:

`/./:`, `/./@:`

Character special and symlink to character special.

`/./;,/./@;`

Block special and symlink to block special.

`/./+,/./@+`

Event port (Solaris, Illumos) and symlink to event port.

All other paths denote a “normal” file.

Note

Executable file have no special indicator here. The command **`ls -F`** would use `*` for files that are executable. This marker is not used here.

See Also

py-treesum, *file format*

Bugs

Version 1 of the file format may change in versions of **`py-treesum`** before version `1.0`; backwards compatibility is not really retained before version `1.0`.

4.2 “treesum” Filter Patterns

Description

Filename matching allows several types of patterns. Each pattern should start with its type specification.

`glob:`

: POSIX *glob()* style, case-sensitive, anchored at the begin and end

`iglob:`

: case-insensitive variant of `glob`

`re:`

: regular expression (Python style)

`path:`

: plain path name (rooted), can be a file or a directory or a prefix thereof

`fullpath:`

: exactly a single full path (file or directory), relative to the root of the tree

If a type is not given explicitely `glob:` is assumed.

Important

- Pattern matching is done using paths *relative* to the root of the directory tree.

  Patterns with absolute paths—while syntactically allowed—never match in *py-treesum*.

- The directory separator is always `/`.

Todo

Encoding used for pattern matching. Also make it more consistent.

Glob Syntax Rules

- The * character matches zero or more characters of a name component without crossing directory boundaries.
- The ** characters matches zero or more characters crossing directory boundaries.
- **/ matches zero or more subdirectories; files do not match.
- The ? character matches exactly one character of a name component.
- The backslash character (\) is used to escape characters that would otherwise be interpreted as special characters. The expression \\ matches a single backslash and \{ matches a left brace for example.
- The [ ] characters are a bracket expression that match a single character of a name component out of a set of characters. For example, [abc] matches a, b, or c. The hyphen (-) may be used to specify a range so [a-z] specifies a range that matches from a to z (inclusive). These forms can be mixed so [abce-g] matches a, b, c, e, f or g.

  If the character after the [is a ! then it is used for negation so [!a-c] matches any character except a, b, or c.

  Within a bracket expression the *, ? and \ characters match themselves.

  The - character matches itself if it is the first or last character within the brackets, or the first or last character after the ! if negating.

  Also, the] character matches itself if it is the first character within the brackets, or the first character after the ! if negating.
- The curly brace characters { } denote a group of subpatterns, where the group matches if any subpattern in the group matches.

  The , character is used to separate the subpatterns. Groups can be nested.
- Leading period/dot characters in file name are treated as regular characters in match operations. For example, the * glob pattern matches file name .login.
- All other characters match themselves.

Examples

`glob:*.py`

: any name ending with .py in the root directory

`*.py`

: the same as glob:*.py (because glob: is the default)

`re:\A[^/]*\.py\Z`

: the same as glob:*.py

`glob:**.py`

any name ending with `.py` anywhere

`re:\.py\Z`

the same as `glob:**.py`

`glob:dir/*`

any name in directory `dir`

Each of these patterns specify any name below directory `dir`:

`glob:dir/**`

`re:\Adir/`

`path:dir/`

Each of these patterns specify any name in any directory that ends with `file`:

`glob:**/file`

`re:(^|/)file\Z`

These patterns specify a single path:

`filepath:dir1/dir2/file`

`re:\Adir1/dir2/file\Z`

History

See also

- https://docs.oracle.com/javase/8/docs/api/java/nio/file/FileSystem.html#getPathMatcher-java.lang.String-
- https://docs.oracle.com/javase/tutorial/essential/io/fileOps.html#glob
- Mercurial[20]: **`hg help patterns`**

See Also

py-treesum, *py-treesum generate*

[20] https://www.mercurial-scm.org/

CHAPTER 5

Implementation Notes

5.1 Important Observations and Facts

5.1.1 Python and Locale Settings

See also

For locale settings:

- https://wiki.archlinux.org/title/Locale
- https://man7.org/linux/man-pages/man7/locale.7.html
- https://man7.org/linux/man-pages/man3/setlocale.3.html
- https://man.freebsd.org/cgi/man.cgi?query=locale&sektion=1&apropos=0&manpath=FreeBSD+15.1-RELEASE+and+Ports.quarterly
- https://man.freebsd.org/cgi/man.cgi?query=setlocale&sektion=3&apropos=0&manpath=FreeBSD+15.1-RELEASE+and+Ports.quarterly
- https://pubs.opengroup.org/onlinepubs/9799919799/basedefs/V1_chap08.html#tag_08

Python:

- https://vstinner.github.io/python3-locales-encodings.html

A locale name generally has the form:

```
language[_territory][.codeset][@modifier]
```

where *language* is an ISO 639 language code, *territory* is an ISO 3166 country code, and *codeset* is a character set or encoding identifier like ISO8859-1 or UTF-8. For a list of all supported locales, try **`locale -a`**.

- The relevant locale setting for the *FSENCODING* is `LC_CTYPE`.
- For the (legacy) plain `C` locale the behaviour in Python[21] has changed over time:
  - Older versions use `US-ASCII` or `ASCII` as the effective filesystem encoding
  - With **PEP 538**[22] (Coercing the legacy C locale to a UTF-8 based locale) Python 3.7+ uses the `UTF-8` encoding instead.
- For Python to recognize a *codeset* a really existing locale must be used: for a list see the output of **`locale -a`**. For example; `en_US.ISO8859-1` or `C.UTF-8` are existing locales on FreeBSD[23], but `C.ISO8859-1` is not a valid locale.

  In this case `sys.getfilesystemencoding()`[24] and other locale dependent things behave as configured and (hopefully) intended by the user.

  Non-existing locales behave differently: for example, on FreeBSD a locale like `C.ISO8859-1` does not exist; it behaves just like plain `C` in Python.
- Existing locales with a *codeset* that is not available in Python may lead to core dumps or other initialization errors.
- Typically, the `POSIX` or `C` locales do not contain as many *codeset* variants as other locales. On FreeBSD there are just `C` and `C.UTF-8`.

Todo

Windows: Is there a command to list all available locales? Registry keys? “NLS” support?

5.1.2 Open Questions

VFAT and Encodings

The handling of mountpoint-specific encoding settings (typical when mounting (Windows) media with VFAT/FAT/exFAT filesystems and others on POSIX systems)

Todo

Mountpoint-specific encodings seem too complicated—maybe later?

[21] https://www.python.org/
[22] https://peps.python.org/pep-0538/
[23] https://www.freebsd.org/
[24] https://docs.python.org/3/library/sys.html#sys.getfilesystemencoding

On FreeBSD for **mount_msdosfs**:

```
Options:

  -L locale
     Specify locale name used for file name conversions
     for DOS and Win'95 names.
     By default ISO 8859-1 assumed as local character
     set.

  -D DOS_codepage
     Specify the MS-DOS code page (aka IBM/OEM code page)
     name used for file name conversions for DOS names.

Examples:

  To mount a Russian MS-DOS file system located in /dev/
↪ada1s1:

    mount_msdosfs -L ru_RU.KOI8-R -D CP866 /dev/ada1s1 /mnt

  To mount a Japanese MS-DOS file system located in /dev/
↪ada1s1:

    mount_msdosfs -L ja_JP.eucJP -D CP932 /dev/ada1s1 /mnt
```

5.2 Testing

Unit tests are to be run from the source distribution.

The most comfortable method is to run the tests via **make**. The provided `Makefile` is a BSD-style makefile. GNU make probably does not work.

> **Note**
>
> For Windows see *this section*.

Targets:

test
: Run all the “normal” unit tests.

  Runs with the user’s current locale; some test scripts are called with some additional locales also.

test-extra
: Run a test script that generates message digests for a file that has an invalid UTF-8 byte sequence. This file is created and removed dynamically and not part of the test data in the source distribution.

This script is run several times with most available locales on the system. Ignored are only codesets that are not available as Python codecs.

test-locale-props
Check some assumptions and observations with regard to Python and locales.

test-all
All of the above tests in a single target

Environment variables that affect tests:

PYTHON
The command to run the Python executable that is used for all tests.

Default: `python`

SKIP-LONG-TESTS
If this variable is set then all test scripts skip long-running tests.

TEST-TMPDIR
This directory will be created and destroyed for each test. Its parent directory must exist already.

Default: `tests/data/_tmp`

5.2.1 Windows

Currently all tests are targeted to POSIX systems. Test data in the `tests/data/` folder contains (POSIX) symbolic links.

Running all or some test scripts on Windows using the **make** targets is not supported.

See also

- https://docs.python.org/3/library/locale.html
- https://vstinner.github.io/python3-locales-encodings.html
- https://en.wikipedia.org/wiki/Locale_(computer_software)
- https://learn.microsoft.com/en-us/windows/win32/Intl/code-page-identifiers

Some single test scripts should work with Windows.

- To test basic assumptions about codepages, locales with Python:

```
setlocal
set LC_ALL=C
python tests/ltest__C.py -v
set LC_ALL=.1252
python tests/ltest__en_US_ISO8859_1.py -v
```

(continues on next page)

(continued from previous page)

```
set LC_ALL=.65001
python tests/ltest__C_UTF_8.py -v
set LC_ALL=.20127
python tests/ltest__en_US_US_ASCII.py -v
endlocal
```

- A subset of “normal” unit tests:

```
python tests/test_shasum.py -v
python tests/test_crc32.py -v
python tests/test_crcmod.py -v
python tests/test_misc.py
python tests/test_match.py -v
python tests/test_encodings.py -v
python tests/test_walk.py -v
python tests/test_treesum.py -v
```

- The special test script with invalid UTF-8 names:

```
setlocal
@REM Using the default locale
python tests/test_treesum_invalid_filenames.py -v
@REM An all other interesting locales
set LC_ALL=.1252
python tests/test_treesum_invalid_filenames.py -v
set LC_ALL=.65001
python tests/test_treesum_invalid_filenames.py -v
set LC_ALL=.20127
python tests/test_treesum_invalid_filenames.py -v
@REM ...
endlocal
```

Warning

These assumptions and script proposals are not yet tested.

Todo

More Windows specific tests

CHAPTER 6

Changes

6.1 Pre-1.0 Series

Development

current (unreleased)

- **[doc]**
  Prepared the first version of a package documentation generated with Sphinx[25] (HTML, PDF (LuaLaTeX) and some manual pages).

py-treesum

- **[breaking]**
  Combine Unix epoch and ISO timestamp values into one record. The parser recognizes the previous format also.
- **[breaking]**
  Option `--ignore-symlink-type` is now the default. Deactivate with `--no-ignore-symlink-type`.
- **[breaking]**
  Shorten some record markers in the "tabular" output style: `ERROR:` into `ERR:` and `WARNING:` into `WARN:`.

  Because no parser existed previously the compatibility break is somewhat limited.

[25] https://sphinx-doc.org/

- **[breaking]**
  In the “tabular” output style errors regarding size numbers are sequences of / characters instead of ? previously.

  > Because no parser existed previously the compatibility break is somewhat limited.

- **[feature]**
  Implemented a parser for the “tabular” output style.

- **[feature]**
  Allow to put warnings into the output. May be dectivated with option `without-warnings`.

- **[feature]**
  Implemented option `--with-notes` to optionally put detailed data about ignored files or directories into the output.

- **[bugfix]**
  In Python 2: the “surrogateescape” error handler (introduced in v0.11) is not called when is should be called: some illegal UTF-8 is accepted while it should be treated as illegal. This is fixed now.

- **[bugfix]**
  Obey `--ignore-symlink-type` also for symbolic links on the command lines.

0.12 (2026-08-29)

py-treesum

- **[feature]**
  Implemented option `--ignore-symlink-type`: when not following symlinks to directories and files do also not try to determine the link target’s type (which can be done only when resolving the link).

0.11.1 (2026-08-28)

- **[bugfix]**
  Fixed wrong publication date in the README.

0.11 (2026-08-28)

py-treesum

- **[breaking]**
  Message digests for directories that contain non-ASCII filenames changed because of a new filename normalization algorithm.

  The digest values are now independent of the `--utf8` or `--fsencoding` settings. These settings now just determine how filenames in the output are encoded.

  These are tested on a POSIX system using Python versions 2.7, 3.4, 3.5, 3.6, 3.7, 3.9, 3.11 and 3.14 using different locale settings.

- **[breaking]**
  On most handled errors do not use the concrete OS *errno* for digesting but an OS independent generic one. Note that this may change message digests in the output.

- **[breaking]**
  Some record formats of a “treesum” file changed to be more consistent.

- **[breaking]**
  Make `--utf8` the default instead of `--fsencoding`.

- **[breaking]**
  Computation of the CRC is now independent of the native OS lineendings that are in effect. Also lines with no records (whitespace only lines, comment lines using `;` or *#*) do not contribute to the CRC.

- **[feature]**
  Implemented `--rfc1924` and `--base85` for shorter digest representations in the output. Also implemented `--base32` for completeness.

- **[bugfix]**
  In some cases the *TAINTED* property of a message digest was not propagated to parent directories properly.

0.10 (2026-08-02)

- **[misc]**
  Modernize the package builds: switched to a **PEP 517**[26] and **PEP 518**[27] compatible build system (using `pyproject.toml`).

  For older Python versions there is a branch `build-2.7` in Mercurial that allows builds with older Python versions (using the “legacy” `setup.py`).

[26] https://peps.python.org/pep-0517/
[27] https://peps.python.org/pep-0518/

0.9 (2025-05-18)

py-treesum

- **[bugfix]**
  Fix an oversight that led to crashes on Python version before 2.7.9 (where `hashlib.algorithms_available`[28] and `hashlib.algorithms_guaranteed`[29] are not available).
- **[bugfix]**
  Optimize CR/LF handling on Windows when using I/O redirection.
- **[misc]**
  Make ERROR outputs more consistent.

0.8.1 (2025-04-02)

This is the first non-private release.

6.2 Breaking Changes

current

- Combine Unix epoch and ISO timestamp values into one record line.
- `--ignore-symlink-type` is now the default.
- Shorten some record markers in the “tabular” output style: `ERROR:` into `ERR:` and `WARNING:` into `WARN:`.
- In the “tabular” output style errors regarding size numbers are sequences of / characters instead of ? previously.

0.11 (2026-08-28)

py-treesum

- Message digests for directories that contain non-ASCII filenames changed because of a new filename normalization algorithm.
- Changed the way how `OSError`[30] exceptions contribute to message digests.
- Some record formats of a “treesum” file changed to be more consistent.

[28] https://docs.python.org/3/library/hashlib.html#hashlib.algorithms_available
[29] https://docs.python.org/3/library/hashlib.html#hashlib.algorithms_guaranteed
[30] https://docs.python.org/3/builtins/exceptions.html#OSError

- Computation of the CRC is now independent of the native OS lineendings that are in effect. Also lines with no records (whitespace only lines, comment lines using ; or #) do not contribute to the CRC.

CHAPTER 7

Licenses

The project is licensed under the terms of the BSD 3-Clause "New" or "Revised" License:

```
Copyright (c) <year> <owner>.

Redistribution and use in source and binary forms, with or without␣
↪modification, are permitted provided that the following conditions are␣
↪met:

1. Redistributions of source code must retain the above copyright notice,␣
↪this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright␣
↪notice, this list of conditions and the following disclaimer in the␣
↪documentation and/or other materials provided with the distribution.

3. Neither the name of the copyright holder nor the names of its␣
↪contributors may be used to endorse or promote products derived from this␣
↪software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS
↪" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO,␣
↪THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR␣
↪PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR␣
↪CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL,␣
↪EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO,␣
↪PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR␣
↪PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF␣
↪LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING␣
↪NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS␣
```

(continues on next page)

(continued from previous page)

```
↪SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
```

LaTeX style files for Sphinx and related files are licensed under the termes of `BSD-3-Clause OR MIT`.

The `MIT License` reads as follows:

```
MIT License

Copyright (c) <year> <copyright holders>

Permission is hereby granted, free of charge, to any person obtaining a␣
↪copy of this software and
associated documentation files (the ”Software”), to deal in the Software␣
↪without restriction, including
without limitation the rights to use, copy, modify, merge, publish,␣
↪distribute, sublicense, and/or sell
copies of the Software, and to permit persons to whom the Software is␣
↪furnished to do so, subject to the
following conditions:

The above copyright notice and this permission notice shall be included␣
↪in all copies or substantial
portions of the Software.

THE SOFTWARE IS PROVIDED ”AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS␣
↪OR IMPLIED, INCLUDING BUT NOT
LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR␣
↪PURPOSE AND NONINFRINGEMENT. IN NO
EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM,␣
↪DAMAGES OR OTHER LIABILITY, WHETHER
IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN␣
↪CONNECTION WITH THE SOFTWARE OR THE
USE OR OTHER DEALINGS IN THE SOFTWARE.
```

Index

Symbols

-A
  py-treesum-generate command line option, 11
-B
  py-shasum command line option, 7
-C
  py-shasum command line option, 8
-E
  py-genpwd command line option, 5
-G
  py-genpwd command line option, 5
-H
  py-treesum-generate command line option, 13
-I
  py-treesum-generate command line option, 14
-L
  py-treesum-generate command line option, 14
-P
  py-treesum-generate command line option, 15
-V
  py-dos2unix command line option, 3
-X
  py-treesum-generate command line option, 12
--accept-treesum
  py-treesum-generate command line option, 11
--algorithm
  py-genpwd command line option, 5
  py-shasum command line option, 7
  py-treesum-generate command line option, 11
--append-output
  py-treesum-generate command line option, 12
--base32
  py-treesum-generate command line option, 12
--base64
  py-shasum command line option, 7
  py-treesum-generate command line option, 12
--base85
  py-treesum-generate command line option, 12
--binary
  py-shasum command line option, 7
--bsd
  py-shasum command line option, 7
  py-treesum-generate command line option, 12
--check
  py-shasum command line option, 7
--checklist
  py-shasum command line option,

8
--checklist-allow-distinfo
py-shasum command line option, 8
--comment
py-treesum-generate command line option, 12
--debug
py-treesum command line option, 9
--exclude
py-treesum-generate command line option, 12
--follow-directory-symlinks
py-treesum-generate command line option, 12
--follow-file-symlinks
py-treesum-generate command line option, 12
--follow-symlinks
py-shasum command line option, 8
--fs-encoding
py-treesum-generate command line option, 12
--fsencoding
py-treesum-generate command line option, 12
--full-mode
py-treesum-generate command line option, 12
--generator
py-treesum-generate command line option, 13
--group
py-genpwd command line option, 5
--group-sep
py-genpwd command line option, 5
--group-size
py-genpwd command line option, 6
--grouping-separator
py-treesum-generate command line option, 13
--half
py-treesum-generate command line option, 13
--help
py-dos2unix command line option, 3
py-genpwd command line option, 5
py-shasum command line option, 7
py-treesum command line option, 9
py-treesum-generate command line option, 11
py-treesum-info command line option, 19
--hex
py-treesum-generate command line option, 13
--ignore-symlink-type
py-treesum-generate command line option, 13
--include
py-treesum-generate command line option, 14
--keepdate
py-dos2unix command line option, 3
--last
py-treesum-info command line option, 19
--logical
py-treesum-generate command line option, 14
--minimal
py-treesum-generate command line option, 14
--mmap
py-shasum command line option, 8
py-treesum-generate command line option, 14
--mode
py-treesum-generate command line option, 14
--mtime
py-treesum-generate command line option, 14
--newfile
py-dos2unix command line op-

tion, 4
--no-follow-directory-symlinks
py-treesum-generate command line option, 14
--no-follow-file-symlinks
py-treesum-generate command line option, 15
--no-ignore-symlink-type
py-treesum-generate command line option, 15
--no-mmap
py-shasum command line option, 8
py-treesum-generate command line option, 15
--oldfile
py-dos2unix command line option, 3
--output
py-treesum-generate command line option, 15
--output-style
py-treesum-generate command line option, 15
--physical
py-treesum-generate command line option, 15
--print-size
py-treesum-generate command line option, 15
--quiet
py-dos2unix command line option, 4
--recurse
py-shasum command line option, 8
--repertoire
py-genpwd command line option, 6
--reverse
py-shasum command line option, 8
py-treesum-generate command line option, 15
--rfc1924
py-treesum-generate command line option, 16
--size-only
py-treesum-generate command line option, 16
--size-width
py-treesum-generate command line option, 16
--tag
py-shasum command line option, 8
py-treesum-generate command line option, 16
--text
py-shasum command line option, 8
--utf8
py-treesum-generate command line option, 16
--utf-8
py-treesum-generate command line option, 16
--version
py-dos2unix command line option, 3
py-genpwd command line option, 5
py-shasum command line option, 7
py-treesum command line option, 9
--with-notes
py-treesum-generate command line option, 16
--without-warnings
py-treesum-generate command line option, 16
-a
py-genpwd command line option, 5
py-shasum command line option, 7
py-treesum-generate command line option, 11
-b
py-shasum command line option, 7
-c
py-shasum command line option, 7
-h

py-dos2unix command line option, 3
py-genpwd command line option, 5
py-shasum command line option, 7
py-treesum command line option, 9
py-treesum-generate command line option, 11
py-treesum-info command line option, 19
-k
py-dos2unix command line option, 3
-l
py-treesum-generate command line option, 12
-n
py-dos2unix command line option, 4
-o
py-dos2unix command line option, 3
py-treesum-generate command line option, 15
-p
py-treesum-generate command line option, 15
-q
py-dos2unix command line option, 4
-r
py-genpwd command line option, 6
py-shasum command line option, 8
py-treesum-generate command line option, 15
-t
py-shasum command line option, 8
-v
py-genpwd command line option, 5
py-shasum command line option, 7
py-treesum command line option, 9

A

ACCEPT-TREESUM
records, 26

B

Block, 21
BSD 3-Clause ”New” or ”Revised” License, 57

C

COMMENT
records, 26
CRC32
records, 24

D

DIGEST
records, 26

E

END-TIMESTAMP
records, 26
environment variable
LC_CTYPE, 46
PYTHON, 48
SKIP-LONG-TESTS, 48
TEST-TMPDIR, 48
ERROR
records, 26

F

file format
treesum, 21
filetypes
py-treesum subcommands, 10
treesum, 37
filter patterns
treesum, 41
FLAGS
records, 24
FNMATCH
records, 26
FSENCODING
records, 24

G

gen

py-treesum subcommands, 9
generate
py-treesum subcommands, 11
GENERATOR
records, 24

H

help
py-treesum subcommands, 9

I

info
py-treesum subcommands, 19

L

LC_CTYPE, 46
licenses, 57
BSD 3-Clause ”New” or ”Revised” License, 57
MIT License, 58

M

MIT License, 58

N

NOTE
records, 26

P

patterns
py-treesum subcommands, 10
treesum filter patterns, 41
py-dos2unix command line option
-V, 3
--help, 3
--keepdate, 3
--newfile, 4
--oldfile, 3
--quiet, 4
--version, 3
-h, 3
-k, 3
-n, 4
-o, 3
-q, 4
py-genpwd command line option
-E, 5
-G, 5
--algorithm, 5
--group, 5
--group-sep, 5
--group-size, 6
--help, 5
--repertoire, 6
--version, 5
-a, 5
-h, 5
-r, 6
-v, 5
py-shasum command line option
-B, 7
-C, 8
--algorithm, 7
--base64, 7
--binary, 7
--bsd, 7
--check, 7
--checklist, 8
--checklist-allow-distinfo, 8
--follow-symlinks, 8
--help, 7
--mmap, 8
--no-mmap, 8
--recurse, 8
--reverse, 8
--tag, 8
--text, 8
--version, 7
-a, 7
-b, 7
-c, 7
-h, 7
-r, 8
-t, 8
-v, 7
py-treesum command line option
--debug, 9
--help, 9
--version, 9
-h, 9
-v, 9
py-treesum subcommands
filetypes, 10
gen, 9
generate, 11
help, 9
info, 19

patterns, 10
version, 9
py-treesum-generate command line option
-A, 11
-H, 13
-I, 14
-L, 14
-P, 15
-X, 12
--accept-treesum, 11
--algorithm, 11
--append-output, 12
--base32, 12
--base64, 12
--base85, 12
--bsd, 12
--comment, 12
--exclude, 12
--follow-directory-symlinks, 12
--follow-file-symlinks, 12
--fs-encoding, 12
--fsencoding, 12
--full-mode, 12
--generator, 13
--grouping-separator, 13
--half, 13
--help, 11
--hex, 13
--ignore-symlink-type, 13
--include, 14
--logical, 14
--minimal, 14
--mmap, 14
--mode, 14
--mtime, 14
--no-follow-directory-symlinks, 14
--no-follow-file-symlinks, 15
--no-ignore-symlink-type, 15
--no-mmap, 15
--output, 15
--output-style, 15
--physical, 15
--print-size, 15
--reverse, 15
--rfc1924, 16
--size-only, 16
--size-width, 16
--tag, 16
--utf8, 16
--utf-8, 16
--with-notes, 16
--without-warnings, 16
-a, 11
-h, 11
-l, 12
-o, 15
-p, 15
-r, 15
py-treesum-info command line option
--help, 19
--last, 19
-h, 19
Python Enhancement Proposals
PEP 517, 53
PEP 518, 53
PEP 528, 10
PEP 529, 10
PEP 538, 46
PEP 540, 10
PEP 686, 10

R

records
ACCEPT-TREESUM, 26
COMMENT, 26
CRC32, 24
DIGEST, 26
END-TIMESTAMP, 26
ERROR, 26
FLAGS, 24
FNMATCH, 26
FSENCODING, 24
GENERATOR, 24
NOTE, 26
ROOT, 26
SIZE, 26
START-TIMESTAMP, 26
VERSION, 24
WARNING, 26
RFC
RFC 1924, 16
ROOT
records, 26

S

SIZE
   records, 26
START-TIMESTAMP
   records, 26

T

treesum
   file format, 21
   filetypes, 37
   filter patterns, 41

V

VERSION
   records, 24
version
   py-treesum subcommands, 9

W

WARNING
   records, 26