

11.9. `shutil` — High-level file operations

Source code: [Lib/shutil.py](#)

The `shutil` module offers a number of high-level operations on files and collections of files. In particular, functions are provided which support file copying and removal. For operations on individual files, see also the `os` module.

Warning: Even the higher-level file copying functions (`shutil.copy()`, `shutil.copy2()`) cannot copy all file metadata.

On POSIX platforms, this means that file owner and group are lost as well as ACLs. On Mac OS, the resource fork and other metadata are not used. This means that resources will be lost and file type and creator codes will not be correct. On Windows, file owners, ACLs and alternate data streams are not copied.

11.9.1. Directory and files operations

`shutil.copyfileobj(fsrc, fdst[, length])` ¶

Copy the contents of the file-like object *fsrc* to the file-like object *fdst*. The integer *length*, if given, is the buffer size. In particular, a negative *length* value means to copy the data without looping over the source data in chunks; by default the data is read in chunks to avoid uncontrolled memory consumption. Note that if the current file position of the *fsrc* object is not 0, only the contents from the current file position to the end of the file will be copied.

`shutil.copyfile(src, dst, *, follow_symlinks=True)`

Copy the contents (no metadata) of the file named *src* to a file named *dst* and return *dst*. *src* and *dst* are path names given as strings. *dst* must be the complete target file name; look at `shutil.copy()` for a copy that accepts a target directory path. If *src* and *dst* specify the same file, `Error` is raised.

The destination location must be writable; otherwise, an `OSError` exception will be raised. If *dst* already exists, it will be replaced. Special files such as character or block devices and pipes cannot be copied with this function.

If *follow_symlinks* is false and *src* is a symbolic link, a new symbolic link will be created instead of copying the file *src* points to.

Changed in version 3.3: `IOError` used to be raised instead of `OSError`. Added *follow_symlinks* argument. Now returns *dst*.

`shutil.copymode(src, dst, *, follow_symlinks=True)`

Copy the permission bits from *src* to *dst*. The file contents, owner, and group are unaffected. *src* and *dst* are path names given as strings. If *follow_symlinks* is false, and both *src* and *dst* are symbolic links, `copymode()` will attempt to modify the mode of *dst* itself (rather than the file it points to). This functionality is not available on every platform; please see `copystat()` for more information. If `copymode()` cannot modify symbolic links on the local platform, and it is asked to do so, it will do nothing and return.

Changed in version 3.3: Added *follow_symlinks* argument.

`shutil.copystat(src, dst, *, follow_symlinks=True)`

Copy the permission bits, last access time, last modification time, and flags from *src* to *dst*. On Linux, `copystat()` also copies the “extended attributes” where possible. The file contents, owner, and group are unaffected. *src* and *dst* are path names given as strings.

If *follow_symlinks* is false, and *src* and *dst* both refer to symbolic links, `copystat()` will operate on the symbolic links themselves rather than the files the symbolic links refer to—reading the information from the *src* symbolic link, and writing the information to the *dst* symbolic link.

Note: Not all platforms provide the ability to examine and modify symbolic links. Python itself can tell you what functionality is locally available.

- If `os.chmod` in `os.supports_follow_symlinks` is True, `copystat()` can modify the permission bits of a symbolic link.
- If `os.utime` in `os.supports_follow_symlinks` is True, `copystat()` can modify the last access and modification times of a symbolic link.
- If `os.chflags` in `os.supports_follow_symlinks` is True, `copystat()` can modify the flags of a symbolic link. (`os.chflags` is not available on all platforms.)

On platforms where some or all of this functionality is unavailable, when asked to modify a symbolic link, `copystat()` will copy everything it can. `copystat()` never returns failure.

Please see `os.supports_follow_symlinks` for more information.

Changed in version 3.3: Added *follow_symlinks* argument and support for Linux extended attributes.

`shutil.copy(src, dst, *, follow_symlinks=True)`

Copies the file *src* to the file or directory *dst*. *src* and *dst* should be strings. If *dst* specifies a directory, the file will be copied into *dst* using the base filename from *src*. Returns the path to the newly created file.

If *follow_symlinks* is false, and *src* is a symbolic link, *dst* will be created as a symbolic link. If *follow_symlinks* is true and *src* is a symbolic link, *dst* will be a copy of the file *src* refers to.

`copy()` copies the file data and the file's permission mode (see `os.chmod()`). Other metadata, like the file's creation and modification times, is not preserved. To preserve all file metadata from the original, use `copy2()` instead.

Changed in version 3.3: Added *follow_symlinks* argument. Now returns path to the newly created file.

`shutil.copy2(src, dst, *, follow_symlinks=True)`

Identical to `copy()` except that `copy2()` also attempts to preserve all file metadata.

When *follow_symlinks* is false, and *src* is a symbolic link, `copy2()` attempts to copy all metadata from the *src* symbolic link to the newly-created *dst* symbolic link. However, this functionality is not available on all platforms. On platforms where some or all of this functionality is unavailable, `copy2()` will preserve all the metadata it can; `copy2()` never returns failure.

`copy2()` uses `copystat()` to copy the file metadata. Please see `copystat()` for more information about platform support for modifying symbolic link metadata.

Changed in version 3.3: Added *follow_symlinks* argument, try to copy extended file system attributes too (currently Linux only). Now returns path to the newly created file.

`shutil.ignore_patterns(*patterns)`

This factory function creates a function that can be used as a callable for `copytree()`'s *ignore* argument, ignoring files and directories that match one of the glob-style *patterns* provided. See the example below.

`shutil.copytree(src, dst, symlinks=False, ignore=None, copy_function=copy2, ignore_dangling_symlinks=False)`

Recursively copy an entire directory tree rooted at *src*, returning the destination directory. The destination directory, named by *dst*, must not already exist; it will be created as well as missing parent directories. Permissions and times of directories are copied with `copystat()`, individual files are copied using `shutil.copy2()`.

If *symlinks* is true, symbolic links in the source tree are represented as symbolic links in the new tree and the metadata of the original links will be copied as far as the platform allows; if false or omitted, the contents and metadata of the linked files are copied to the new tree.

When *symlinks* is false, if the file pointed by the symlink doesn't exist, a exception will be added in the list of errors raised in a `Error` exception at the end of the copy

process. You can set the optional *ignore_dangling_symlinks* flag to true if you want to silence this exception. Notice that this option has no effect on platforms that don't support `os.symlink()`.

If *ignore* is given, it must be a callable that will receive as its arguments the directory being visited by `copytree()`, and a list of its contents, as returned by `os.listdir()`. Since `copytree()` is called recursively, the *ignore* callable will be called once for each directory that is copied. The callable must return a sequence of directory and file names relative to the current directory (i.e. a subset of the items in its second argument); these names will then be ignored in the copy process. `ignore_patterns()` can be used to create such a callable that ignores names based on glob-style patterns.

If exception(s) occur, an `Error` is raised with a list of reasons.

If *copy_function* is given, it must be a callable that will be used to copy each file. It will be called with the source path and the destination path as arguments. By default, `shutil.copy2()` is used, but any function that supports the same signature (like `shutil.copy()`) can be used.

Changed in version 3.3: Copy metadata when *symlinks* is false. Now returns *dst*.

Changed in version 3.2: Added the *copy_function* argument to be able to provide a custom copy function. Added the *ignore_dangling_symlinks* argument to silent dangling symlinks errors when *symlinks* is false.

`shutil.rmtree(path, ignore_errors=False, onerror=None)`

Delete an entire directory tree; *path* must point to a directory (but not a symbolic link to a directory). If *ignore_errors* is true, errors resulting from failed removals will be ignored; if false or omitted, such errors are handled by calling a handler specified by *onerror* or, if that is omitted, they raise an exception.

Note: On platforms that support the necessary fd-based functions a symlink attack resistant version of `rmtree()` is used by default. On other platforms, the `rmtree()` implementation is susceptible to a symlink attack: given proper timing and circumstances, attackers can manipulate symlinks on the filesystem to delete files they wouldn't be able to access otherwise. Applications can use the `rmtree.avoids_symlink_attacks` function attribute to determine which case applies.

If *onerror* is provided, it must be a callable that accepts three parameters: *function*, *path*, and *excinfo*.

The first parameter, *function*, is the function which raised the exception; it depends on the platform and implementation. The second parameter, *path*, will be the path name

passed to *function*. The third parameter, *excinfo*, will be the exception information returned by `sys.exc_info()`. Exceptions raised by *onerror* will not be caught.

Changed in version 3.3: Added a symlink attack resistant version that is used automatically if platform supports fd-based functions.

`rmtree.avoids_symlink_attacks`

Indicates whether the current platform and implementation provides a symlink attack resistant version of `rmtree()`. Currently this is only true for platforms supporting fd-based directory access functions.

New in version 3.3.

`shutil.move(src, dst)`

Recursively move a file or directory (*src*) to another location (*dst*) and return the destination.

If the destination is a directory or a symlink to a directory, then *src* is moved inside that directory.

The destination directory must not already exist. If the destination already exists but is not a directory, it may be overwritten depending on `os.rename()` semantics.

If the destination is on the current filesystem, then `os.rename()` is used. Otherwise, *src* is copied (using `shutil.copy2()`) to *dst* and then removed. In case of symlinks, a new symlink pointing to the target of *src* will be created in or as *dst* and *src* will be removed.

Changed in version 3.3: Added explicit symlink handling for foreign filesystems, thus adapting it to the behavior of GNU's **mv**. Now returns *dst*.

`shutil.disk_usage(path)`

Return disk usage statistics about the given path as a *named tuple* with the attributes *total*, *used* and *free*, which are the amount of total, used and free space, in bytes.

New in version 3.3.

Availability: Unix, Windows.

`shutil.chown(path, user=None, group=None)`

Change owner *user* and/or *group* of the given *path*.

user can be a system user name or a uid; the same applies to *group*. At least one argument is required.

See also `os.chown()`, the underlying function.

Availability: Unix.

New in version 3.3.

`shutil.which(cmd, mode=os.F_OK | os.X_OK, path=None)`

Return the path to an executable which would be run if the given *cmd* was called. If no *cmd* would be called, return `None`.

mode is a permission mask passed a to `os.access()`, by default determining if the file exists and executable.

When no *path* is specified, the results of `os.environ()` are used, returning either the “PATH” value or a fallback of `os.defpath`.

On Windows, the current directory is always prepended to the *path* whether or not you use the default or provide your own, which is the behavior the command shell uses when finding executables. Additionally, when finding the *cmd* in the *path*, the `PATHEXT` environment variable is checked. For example, if you call `shutil.which("python")`, `which()` will search `PATHEXT` to know that it should look for `python.exe` within the *path* directories. For example, on Windows:

```
>>> shutil.which("python")
'C:\\Python33\\python.EXE'
```

```
>>>
```

New in version 3.3.

exception `shutil.Error`

This exception collects exceptions that are raised during a multi-file operation. For `copytree()`, the exception argument is a list of 3-tuples (*srcname*, *dstname*, *exception*).

11.9.1.1. copytree example

This example is the implementation of the `copytree()` function, described above, with the docstring omitted. It demonstrates many of the other functions provided by this module.

```
def copytree(src, dst, symlinks=False):
    names = os.listdir(src)
    os.makedirs(dst)
    errors = []
    for name in names:
        srcname = os.path.join(src, name)
        dstname = os.path.join(dst, name)
        try:
            if symlinks and os.path.islink(srcname):
                linkto = os.readlink(srcname)
```

```

        os.symlink(linkto, dstname)
    elif os.path.isdir(srcname):
        copytree(srcname, dstname, symlinks)
    else:
        copy2(srcname, dstname)
        # XXX What about devices, sockets etc.?
    except OSError as why:
        errors.append((srcname, dstname, str(why)))
        # catch the Error from the recursive copytree so that we can
        # continue with other files
    except Error as err:
        errors.extend(err.args[0])

try:
    copystat(src, dst)
except WindowsError:
    # can't copy file access times on Windows
    pass
except OSError as why:
    errors.extend((src, dst, str(why)))
if errors:
    raise Error(errors)

```

Another example that uses the `ignore_patterns()` helper:

```

from shutil import copytree, ignore_patterns

copytree(source, destination, ignore=ignore_patterns('*.pyc', 'tmp*'))

```

This will copy everything except `.pyc` files and files or directories whose name starts with `tmp`.

Another example that uses the `ignore` argument to add a logging call:

```

from shutil import copytree
import logging

def _logpath(path, names):
    logging.info('Working in %s' % path)
    return [] # nothing will be ignored

copytree(source, destination, ignore=_logpath)

```

11.9.2. Archiving operations

New in version 3.2.

High-level utilities to create and read compressed and archived files are also provided.

They rely on the `zipfile` and `tarfile` modules.

```
shutil.make_archive(base_name, format[, root_dir[, base_dir[, verbose[,  
dry_run[, owner[, group[, logger]]]]]])
```

Create an archive file (such as zip or tar) and return its name.

base_name is the name of the file to create, including the path, minus any format-specific extension. *format* is the archive format: one of “zip”, “tar”, “bztar” (if the `bz2` module is available) or “gztar”.

root_dir is a directory that will be the root directory of the archive; for example, we typically `chdir` into *root_dir* before creating the archive.

base_dir is the directory where we start archiving from; i.e. *base_dir* will be the common prefix of all files and directories in the archive.

root_dir and *base_dir* both default to the current directory.

owner and *group* are used when creating a tar archive. By default, uses the current owner and group.

logger must be an object compatible with **PEP 282**, usually an instance of `logging.Logger`.

```
shutil.get_archive_formats()
```

Return a list of supported formats for archiving. Each element of the returned sequence is a tuple (name, description)

By default `shutil` provides these formats:

- *gztar*: gzip’ed tar-file
- *bztar*: bzip2’ed tar-file (if the `bz2` module is available.)
- *tar*: uncompressed tar file
- *zip*: ZIP file

You can register new formats or provide your own archiver for any existing formats, by using `register_archive_format()`.

```
shutil.register_archive_format(name, function[, extra_args[,  
description]])
```

Register an archiver for the format *name*. *function* is a callable that will be used to invoke the archiver.

If given, *extra_args* is a sequence of (name, value) pairs that will be used as extra keywords arguments when the archiver callable is used.

description is used by `get_archive_formats()` which returns the list of archivers.

Defaults to an empty list.

```
shutil.unregister_archive_format(name)
```

Remove the archive format *name* from the list of supported formats.

```
shutil.unpack_archive(filename[, extract_dir[, format]])
```

Unpack an archive. *filename* is the full path of the archive.

extract_dir is the name of the target directory where the archive is unpacked. If not provided, the current working directory is used.

format is the archive format: one of “zip”, “tar”, or “gztar”. Or any other format registered with `register_unpack_format()`. If not provided, `unpack_archive()` will use the archive file name extension and see if an unpacker was registered for that extension. In case none is found, a `ValueError` is raised.

```
shutil.register_unpack_format(name, extensions, function[, extra_args[, description]])
```

Registers an unpack format. *name* is the name of the format and *extensions* is a list of extensions corresponding to the format, like `.zip` for Zip files.

function is the callable that will be used to unpack archives. The callable will receive the path of the archive, followed by the directory the archive must be extracted to.

When provided, *extra_args* is a sequence of (name, value) tuples that will be passed as keywords arguments to the callable.

description can be provided to describe the format, and will be returned by the `get_unpack_formats()` function.

```
shutil.unregister_unpack_format(name)
```

Unregister an unpack format. *name* is the name of the format.

```
shutil.get_unpack_formats()
```

Return a list of all registered formats for unpacking. Each element of the returned sequence is a tuple (name, extensions, description).

By default `shutil` provides these formats:

- *gztar*: gzip'ed tar-file
- *bztar*: bzip2'ed tar-file (if the `bz2` module is available.)
- *tar*: uncompressed tar file
- *zip*: ZIP file

You can register new formats or provide your own unpacker for any existing formats, by using `register_unpack_format()`.

11.9.2.1. Archiving example

In this example, we create a gzip'ed tar-file archive containing all files found in the `.ssh` directory of the user:

```
>>> from shutil import make_archive
>>> import os
>>> archive_name = os.path.expanduser(os.path.join('~', 'myarchive'))
>>> root_dir = os.path.expanduser(os.path.join('~', '.ssh'))
>>> make_archive(archive_name, 'gztar', root_dir)
'/Users/tarek/myarchive.tar.gz'
```

The resulting archive contains:

```
$ tar -tzvf /Users/tarek/myarchive.tar.gz
drwx----- tarek/staff      0 2010-02-01 16:23:40 ./
-rw-r--r-- tarek/staff    609 2008-06-09 13:26:54 ./authorized_keys
-rwxr-xr-x tarek/staff     65 2008-06-09 13:26:54 ./config
-rwx----- tarek/staff    668 2008-06-09 13:26:54 ./id_dsa
-rwxr-xr-x tarek/staff    609 2008-06-09 13:26:54 ./id_dsa.pub
-rw----- tarek/staff   1675 2008-06-09 13:26:54 ./id_rsa
-rw-r--r-- tarek/staff    397 2008-06-09 13:26:54 ./id_rsa.pub
-rw-r--r-- tarek/staff  37192 2010-02-06 18:23:10 ./known_hosts
```

11.9.3. Querying the size of the output terminal

New in version 3.3.

`shutil.get_terminal_size(fallback=(columns, lines))`

Get the size of the terminal window.

For each of the two dimensions, the environment variable, `COLUMNS` and `LINES` respectively, is checked. If the variable is defined and the value is a positive integer, it is used.

When `COLUMNS` or `LINES` is not defined, which is the common case, the terminal connected to `sys.__stdout__` is queried by invoking `os.get_terminal_size()`.

If the terminal size cannot be successfully queried, either because the system doesn't support querying, or because we are not connected to a terminal, the value given in `fallback` parameter is used. `fallback` defaults to `(80, 24)` which is the default size used by many terminal emulators.

The value returned is a named tuple of type `os.terminal_size`.

See also: The Single UNIX Specification, Version 2, [Other Environment Variables](#).