2. Writing the Setup Script

The setup script is the centre of all activity in building, distributing, and installing modules using the Distutils. The main purpose of the setup script is to describe your module distribution to the Distutils, so that the various commands that operate on your modules do the right thing. As we saw in section A Simple Example above, the setup script consists mainly of a call to `setup()`, and most information supplied to the Distutils by the module developer is supplied as keyword arguments to `setup()`.

Here's a slightly more involved example, which we'll follow for the next couple of sections: the Distutils' own setup script. (Keep in mind that although the Distutils are included with Python 1.6 and later, they also have an independent existence so that Python 1.5.2 users can use them to install other module distributions. The Distutils' own setup script, shown here, is used to install the package into Python 1.5.2.)

```
#!/usr/bin/env python

from distutils.core import setup

setup(name='Distutils',
      version='1.0',
      description='Python Distribution Utilities',
      author='Greg Ward',
      author_email='gward@python.net',
      url='https://www.python.org/sigs/distutils-sig/',
      packages=['distutils', 'distutils.command'],
     )
```

There are only two differences between this and the trivial one-file distribution presented in section A Simple Example: more metadata, and the specification of pure Python modules by package, rather than by module. This is important since the Distutils consist of a couple of dozen modules split into (so far) two packages; an explicit list of every module would be tedious to generate and difficult to maintain. For more information on the additional meta-data, see section Additional meta-data.

Note that any pathnames (files or directories) supplied in the setup script should be written using the Unix convention, i.e. slash-separated. The Distutils will take care of converting this platform-neutral representation into whatever is appropriate on your current platform before actually using the pathname. This makes your setup script portable across operating systems, which of course is one of the major goals of the Distutils. In this spirit, all pathnames in this document are slash-separated.

This, of course, only applies to pathnames given to Distutils functions. If you, for example, use standard Python functions such as `glob.glob()` or `os.listdir()` to

specify files, you should be careful to write portable code instead of hardcoding path separators:

```
glob.glob(os.path.join('mydir', 'subdir', '*.html'))
os.listdir(os.path.join('mydir', 'subdir'))
```

2.1. Listing whole packages

The `packages` option tells the Distutils to process (build, distribute, install, etc.) all pure Python modules found in each package mentioned in the `packages` list. In order to do this, of course, there has to be a correspondence between package names and directories in the filesystem. The default correspondence is the most obvious one, i.e. package `distutils` is found in the directory `distutils` relative to the distribution root. Thus, when you say `packages = ['foo']` in your setup script, you are promising that the Distutils will find a file `foo/__init__.py` (which might be spelled differently on your system, but you get the idea) relative to the directory where your setup script lives. If you break this promise, the Distutils will issue a warning but still process the broken package anyway.

If you use a different convention to lay out your source directory, that's no problem: you just have to supply the `package_dir` option to tell the Distutils about your convention. For example, say you keep all Python source under `lib`, so that modules in the "root package" (i.e., not in any package at all) are in `lib`, modules in the `foo` package are in `lib/foo`, and so forth. Then you would put

```
package_dir = {'': 'lib'}
```

in your setup script. The keys to this dictionary are package names, and an empty package name stands for the root package. The values are directory names relative to your distribution root. In this case, when you say `packages = ['foo']`, you are promising that the file `lib/foo/__init__.py` exists.

Another possible convention is to put the `foo` package right in `lib`, the `foo.bar` package in `lib/bar`, etc. This would be written in the setup script as

```
package_dir = {'foo': 'lib'}
```

A `package: dir` entry in the `package_dir` dictionary implicitly applies to all packages below *package*, so the `foo.bar` case is automatically handled here. In this example, having `packages = ['foo', 'foo.bar']` tells the Distutils to look for `lib/__init__.py` and `lib/bar/__init__.py`. (Keep in mind that although `package_dir` applies recursively, you must explicitly list all packages in `packages`:

the Distutils will *not* recursively scan your source tree looking for any directory with an `__init__.py` file.)

2.2. Listing individual modules

For a small module distribution, you might prefer to list all modules rather than listing packages—especially the case of a single module that goes in the "root package" (i.e., no package at all). This simplest case was shown in section A Simple Example; here is a slightly more involved example:

```
py_modules = ['mod1', 'pkg.mod2']
```

This describes two modules, one of them in the "root" package, the other in the `pkg` package. Again, the default package/directory layout implies that these two modules can be found in `mod1.py` and `pkg/mod2.py`, and that `pkg/__init__.py` exists as well. And again, you can override the package/directory correspondence using the `package_dir` option.

2.3. Describing extension modules

Just as writing Python extension modules is a bit more complicated than writing pure Python modules, describing them to the Distutils is a bit more complicated. Unlike pure modules, it's not enough just to list modules or packages and expect the Distutils to go out and find the right files; you have to specify the extension name, source file(s), and any compile/link requirements (include directories, libraries to link with, etc.).

All of this is done through another keyword argument to `setup()`, the `ext_modules` option. `ext_modules` is just a list of `Extension` instances, each of which describes a single extension module. Suppose your distribution includes a single extension, called `foo` and implemented by `foo.c`. If no additional instructions to the compiler/linker are needed, describing this extension is quite simple:

```
Extension('foo', ['foo.c'])
```

The `Extension` class can be imported from `distutils.core` along with `setup()`. Thus, the setup script for a module distribution that contains only this one extension and nothing else might be:

```
from distutils.core import setup, Extension
setup(name='foo',
      version='1.0',
```

```
      ext_modules=[Extension('foo', ['foo.c'])],
      )
```

The `Extension` class (actually, the underlying extension-building machinery implemented by the **build_ext** command) supports a great deal of flexibility in describing Python extensions, which is explained in the following sections.

2.3.1. Extension names and packages

The first argument to the `Extension` constructor is always the name of the extension, including any package names. For example,

```
Extension('foo', ['src/foo1.c', 'src/foo2.c'])
```

describes an extension that lives in the root package, while

```
Extension('pkg.foo', ['src/foo1.c', 'src/foo2.c'])
```

describes the same extension in the `pkg` package. The source files and resulting object code are identical in both cases; the only difference is where in the filesystem (and therefore where in Python's namespace hierarchy) the resulting extension lives.

If you have a number of extensions all in the same package (or all under the same base package), use the `ext_package` keyword argument to `setup()`. For example,

```
setup(...,
      ext_package='pkg',
      ext_modules=[Extension('foo', ['foo.c']),
                   Extension('subpkg.bar', ['bar.c'])],
     )
```

will compile `foo.c` to the extension `pkg.foo`, and `bar.c` to `pkg.subpkg.bar`.

2.3.2. Extension source files

The second argument to the `Extension` constructor is a list of source files. Since the Distutils currently only support C, C++, and Objective-C extensions, these are normally C/C++/Objective-C source files. (Be sure to use appropriate extensions to distinguish C++ source files: `.cc` and `.cpp` seem to be recognized by both Unix and Windows compilers.)

However, you can also include SWIG interface (`.i`) files in the list; the **build_ext** command knows how to deal with SWIG extensions: it will run SWIG on the interface file and compile the resulting C/C++ file into your extension.

This warning notwithstanding, options to SWIG can be currently passed like this:

```
setup(...,
      ext_modules=[Extension('_foo', ['foo.i'],
                             swig_opts=['-modern', '-I../include'])],
      py_modules=['foo'],
     )
```

Or on the commandline like this:

```
> python setup.py build_ext --swig-opts="-modern -I../include"
```

On some platforms, you can include non-source files that are processed by the compiler and included in your extension. Currently, this just means Windows message text (`.mc`) files and resource definition (`.rc`) files for Visual C++. These will be compiled to binary resource (`.res`) files and linked into the executable.

2.3.3. Preprocessor options

Three optional arguments to `Extension` will help if you need to specify include directories to search or preprocessor macros to define/undefine: `include_dirs`, `define_macros`, and `undef_macros`.

For example, if your extension requires header files in the `include` directory under your distribution root, use the `include_dirs` option:

```
Extension('foo', ['foo.c'], include_dirs=['include'])
```

You can specify absolute directories there; if you know that your extension will only be built on Unix systems with X11R6 installed to `/usr`, you can get away with

```
Extension('foo', ['foo.c'], include_dirs=['/usr/include/X11'])
```

You should avoid this sort of non-portable usage if you plan to distribute your code: it's probably better to write C code like

```
#include <X11/Xlib.h>
```

If you need to include header files from some other Python extension, you can take advantage of the fact that header files are installed in a consistent way by the Distutils **install_headers** command. For example, the Numerical Python header files are installed (on a standard Unix installation) to `/usr/local/include/python1.5/Numerical`. (The exact location will differ according to your platform and Python installation.) Since the Python include directo-

ry—`/usr/local/include/python1.5` in this case—is always included in the search path when building Python extensions, the best approach is to write C code like

```
#include <Numerical/arrayobject.h>
```

If you must put the `Numerical` include directory right into your header search path, though, you can find that directory using the Distutils `distutils.sysconfig` module:

```
from distutils.sysconfig import get_python_inc
incdir = os.path.join(get_python_inc(plat_specific=1), 'Numerical')
setup(...,
      Extension(..., include_dirs=[incdir]),
      )
```

Even though this is quite portable—it will work on any Python installation, regardless of platform—it's probably easier to just write your C code in the sensible way.

You can define and undefine pre-processor macros with the `define_macros` and `undef_macros` options. `define_macros` takes a list of `(name, value)` tuples, where `name` is the name of the macro to define (a string) and `value` is its value: either a string or `None`. (Defining a macro `FOO` to `None` is the equivalent of a bare `#define FOO` in your C source: with most compilers, this sets `FOO` to the string `1`.) `undef_macros` is just a list of macros to undefine.

For example:

```
Extension(...,
          define_macros=[('NDEBUG', '1'),
                         ('HAVE_STRFTIME', None)],
          undef_macros=['HAVE_FOO', 'HAVE_BAR'])
```

is the equivalent of having this at the top of every C source file:

```
#define NDEBUG 1
#define HAVE_STRFTIME
#undef HAVE_FOO
#undef HAVE_BAR
```

2.3.4. Library options

You can also specify the libraries to link against when building your extension, and the directories to search for those libraries. The `libraries` option is a list of libraries to link against, `library_dirs` is a list of directories to search for libraries at link-

time, and `runtime_library_dirs` is a list of directories to search for shared (dynamically loaded) libraries at run-time.

For example, if you need to link against libraries known to be in the standard library search path on target systems

```
Extension(...,
          libraries=['gdbm', 'readline'])
```

If you need to link with libraries in a non-standard location, you'll have to include the location in `library_dirs`:

```
Extension(...,
          library_dirs=['/usr/X11R6/lib'],
          libraries=['X11', 'Xt'])
```

(Again, this sort of non-portable construct should be avoided if you intend to distribute your code.)

2.3.5. Other options

There are still some other options which can be used to handle special cases.

The `optional` option is a boolean; if it is true, a build failure in the extension will not abort the build process, but instead simply not install the failing extension.

The `extra_objects` option is a list of object files to be passed to the linker. These files must not have extensions, as the default extension for the compiler is used.

`extra_compile_args` and `extra_link_args` can be used to specify additional command line options for the respective compiler and linker command lines.

`export_symbols` is only useful on Windows. It can contain a list of symbols (functions or variables) to be exported. This option is not needed when building compiled extensions: Distutils will automatically add `initmodule` to the list of exported symbols.

The `depends` option is a list of files that the extension depends on (for example header files). The build command will call the compiler on the sources to rebuild extension if any on this files has been modified since the previous build.

2.4. Relationships between Distributions and Packages

A distribution may relate to packages in three specific ways:

1. It can require packages or modules.
2. It can provide packages or modules.
3. It can obsolete packages or modules.

These relationships can be specified using keyword arguments to the `distutils.core.setup()` function.

Dependencies on other Python modules and packages can be specified by supplying the *requires* keyword argument to `setup()`. The value must be a list of strings. Each string specifies a package that is required, and optionally what versions are sufficient.

To specify that any version of a module or package is required, the string should consist entirely of the module or package name. Examples include `'mymodule'` and `'xml.parsers.expat'`.

If specific versions are required, a sequence of qualifiers can be supplied in parentheses. Each qualifier may consist of a comparison operator and a version number. The accepted comparison operators are:

```
<     >     ==
<=    >=    !=
```

These can be combined by using multiple qualifiers separated by commas (and optional whitespace). In this case, all of the qualifiers must be matched; a logical AND is used to combine the evaluations.

Let's look at a bunch of examples:

Requires Expression	Explanation
`==1.0`	Only version `1.0` is compatible
`>1.0, !=1.5.1, <2.0`	Any version after `1.0` and before `2.0` is compatible, except `1.5.1`

Now that we can specify dependencies, we also need to be able to specify what we provide that other distributions can require. This is done using the *provides* keyword argument to `setup()`. The value for this keyword is a list of strings, each of which names a Python module or package, and optionally identifies the version. If the version is not specified, it is assumed to match that of the distribution.

Some examples:

Provides Expression	Explanation
`mypkg`	Provide `mypkg`, using the distribution version
`mypkg (1.1)`	Provide `mypkg` version 1.1, regardless of the distribution version

A package can declare that it obsoletes other packages using the *obsoletes* keyword argument. The value for this is similar to that of the *requires* keyword: a list of strings giving module or package specifiers. Each specifier consists of a module or package name optionally followed by one or more version qualifiers. Version qualifiers are given in parentheses after the module or package name.

The versions identified by the qualifiers are those that are obsoleted by the distribution being described. If no qualifiers are given, all versions of the named module or package are understood to be obsoleted.

2.5. Installing Scripts

So far we have been dealing with pure and non-pure Python modules, which are usually not run by themselves but imported by scripts.

Scripts are files containing Python source code, intended to be started from the command line. Scripts don't require Distutils to do anything very complicated. The only clever feature is that if the first line of the script starts with `#!` and contains the word "python", the Distutils will adjust the first line to refer to the current interpreter location. By default, it is replaced with the current interpreter location. The `--executable` (or `-e`) option will allow the interpreter path to be explicitly overridden.

The `scripts` option simply is a list of files to be handled in this way. From the PyXML setup script:

```
setup(...,
      scripts=['scripts/xmlproc_parse', 'scripts/xmlproc_val']
      )
```

Changed in version 3.1: All the scripts will also be added to the `MANIFEST` file if no template is provided. See Specifying the files to distribute.

2.6. Installing Package Data

Often, additional files need to be installed into a package. These files are often data that's closely related to the package's implementation, or text files containing docu-

mentation that might be of interest to programmers using the package. These files are called *package data*.

Package data can be added to packages using the `package_data` keyword argument to the `setup()` function. The value must be a mapping from package name to a list of relative path names that should be copied into the package. The paths are interpreted as relative to the directory containing the package (information from the `package_dir` mapping is used if appropriate); that is, the files are expected to be part of the package in the source directories. They may contain glob patterns as well.

The path names may contain directory portions; any necessary directories will be created in the installation.

For example, if a package should contain a subdirectory with several data files, the files can be arranged like this in the source tree:

```
setup.py
src/
    mypkg/
        __init__.py
        module.py
        data/
            tables.dat
            spoons.dat
            forks.dat
```

The corresponding call to `setup()` might be:

```
setup(...,
      packages=['mypkg'],
      package_dir={'mypkg': 'src/mypkg'},
      package_data={'mypkg': ['data/*.dat']},
      )
```

Changed in version 3.1: All the files that match `package_data` will be added to the `MANIFEST` file if no template is provided. See Specifying the files to distribute.

2.7. Installing Additional Files

The `data_files` option can be used to specify additional files needed by the module distribution: configuration files, message catalogs, data files, anything which doesn't fit in the previous categories.

`data_files` specifies a sequence of (*directory*, *files*) pairs in the following way:

```
setup(...,
      data_files=[('bitmaps', ['bm/b1.gif', 'bm/b2.gif']),
                  ('config', ['cfg/data.cfg']),
                  ('/etc/init.d', ['init-script'])]
     )
```

Note that you can specify the directory names where the data files will be installed, but you cannot rename the data files themselves.

Each (*directory*, *files*) pair in the sequence specifies the installation directory and the files to install there. If *directory* is a relative path, it is interpreted relative to the installation prefix (Python's `sys.prefix` for pure-Python packages, `sys.exec_prefix` for packages that contain extension modules). Each file name in *files* is interpreted relative to the `setup.py` script at the top of the package source distribution. No directory information from *files* is used to determine the final location of the installed file; only the name of the file is used.

You can specify the `data_files` options as a simple sequence of files without specifying a target directory, but this is not recommended, and the **install** command will print a warning in this case. To install data files directly in the target directory, an empty string should be given as the directory.

Changed in version 3.1: All the files that match `data_files` will be added to the `MANIFEST` file if no template is provided. See Specifying the files to distribute.

2.8. Additional meta-data

The setup script may include additional meta-data beyond the name and version. This information includes:

Meta-Data	Description	Value	Notes
`name`	name of the package	short string	(1)
`version`	version of this release	short string	(1)(2)
`author`	package author's name	short string	(3)
`author_email`	email address of the package author	email address	(3)
`maintainer`	package maintainer's name	short string	(3)
`maintainer_email`	email address of the package maintainer	email address	(3)
`url`	home page for the package	URL	(1)
`description`	short, summary description of the package	short string	

Meta-Data	Description	Value	Notes
`long_description`	longer description of the package	long string	(5)
`download_url`	location where the package may be downloaded	URL	(4)
`classifiers`	a list of classifiers	list of strings	(4)
`platforms`	a list of platforms	list of strings	
`license`	license for the package	short string	(6)

Notes:

1. These fields are required.
2. It is recommended that versions take the form *major.minor[.patch[.sub]]*.
3. Either the author or the maintainer must be identified. If maintainer is provided, distutils lists it as the author in `PKG-INFO`.
4. These fields should not be used if your package is to be compatible with Python versions prior to 2.2.3 or 2.3. The list is available from the PyPI website.
5. The `long_description` field is used by PyPI when you are registering a package, to build its home page.
6. The `license` field is a text indicating the license covering the package where the license is not a selection from the "License" Trove classifiers. See the `Classifier` field. Notice that there's a `licence` distribution option which is deprecated but still acts as an alias for `license`.

'short string'
: A single line of text, not more than 200 characters.

'long string'
: Multiple lines of plain text in reStructuredText format (see http://docutils.sourceforge.net/).

'list of strings'
: See below.

Encoding the version information is an art in itself. Python packages generally adhere to the version format *major.minor[.patch][sub]*. The major number is 0 for initial, experimental releases of software. It is incremented for releases that represent major milestones in a package. The minor number is incremented when important new features are added to the package. The patch number increments when bug-fix releases are made. Additional trailing version information is sometimes used to indicate sub-releases. These are "a1,a2,…,aN" (for alpha releases, where functionality and API may change), "b1,b2,…,bN" (for beta releases, which only fix bugs) and "pr1,pr2,…,prN" (for final pre-release release testing). Some examples:

0.1.0

the first, experimental release of a package

1.0.1a2
: the second alpha release of the first patch version of 1.0

`classifiers` are specified in a Python list:

```
setup(...,
      classifiers=[
          'Development Status :: 4 - Beta',
          'Environment :: Console',
          'Environment :: Web Environment',
          'Intended Audience :: End Users/Desktop',
          'Intended Audience :: Developers',
          'Intended Audience :: System Administrators',
          'License :: OSI Approved :: Python Software Foundation Licer
          'Operating System :: MacOS :: MacOS X',
          'Operating System :: Microsoft :: Windows',
          'Operating System :: POSIX',
          'Programming Language :: Python',
          'Topic :: Communications :: Email',
          'Topic :: Office/Business',
          'Topic :: Software Development :: Bug Tracking',
          ],
      )
```

2.9. Debugging the setup script

Sometimes things go wrong, and the setup script doesn't do what the developer wants.

Distutils catches any exceptions when running the setup script, and print a simple error message before the script is terminated. The motivation for this behaviour is to not confuse administrators who don't know much about Python and are trying to install a package. If they get a big long traceback from deep inside the guts of Distutils, they may think the package or the Python installation is broken because they don't read all the way down to the bottom and see that it's a permission problem.

On the other hand, this doesn't help the developer to find the cause of the failure. For this purpose, the `DISTUTILS_DEBUG` environment variable can be set to anything except an empty string, and distutils will now print detailed information about what it is doing, dump the full traceback when an exception occurs, and print the whole command line when an external program (like a C compiler) fails.