Python Zip File

Gaurav Kr. suman

MAT12

Python zip file

ZIP is an archive file format that supports lossless data compression. By lossless compression, we mean that the compression algorithm allows the original data to be perfectly reconstructed from the compressed data. So, a ZIP file is a single file containing one or more compressed files, offering an ideal way to make large files smaller and keep related files together.

Why do we need zip files?

- To reduce storage requirements.
- To improve transfer speed over standard connections.

To work on zip files using python, we will use an inbuilt python module called zipfile.

Extracting a zip file

```
# importing required modules
from zipfile import ZipFile

# specifying the zip file name
file_name = "my_python_files.zip"

# opening the zip file in READ mode
with ZipFile(file_name, 'r') as zip:
    # printing all the contents of the zip file
    zip.printdir()

    # extracting all the files
    print('Extracting all the files now...')
    zip.extractall()
    print('Done!')
```

The above program extracts a zip file named “my_python_files.zip” in the same directory as of this python script.
The output of above program may look like this:

```
File Name                                             Modified             Size
python_files/python_basic.pdf                 2016-12-08 23:05:34       210098
python_files/python_wiki.txt                  2016-12-08 23:02:40          990
python_files/logos/python_logo.jpg            2016-12-08 23:00:42        46623

Extracting all the files now...
Done!
```

Let us try to understand the above code in pieces:

- `from zipfile import ZipFile`

ZipFile is a class of zipfile module for reading and writing zip files. Here we import only class ZipFile from zipfile module.

- `with ZipFile(file_name, 'r') as zip:`

  Here, a ZipFile object is made by calling ZipFile constructor which accepts zip file name and mode parameters. We create a ZipFile object in **READ** mode and name it as **zip**.
- `zip.printdir()`

  printdir() method prints a table of contents for the archive.
- `zip.extractall()`

  extractall() method will extract all the contents of the zip file to the current working directory. You can also call **extract()** method to extract any file by specifying its path in the zip file.
  For example:
- ```
  zip.extract('python_files/python_wiki.txt')
  ```

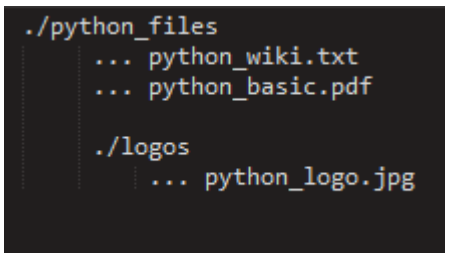
  This will extract only the specified file.

  If you want to read some specific file, you can go like this:

```
data = zip.read(name_of_file_to_read)
```

Writing to a zip file

Consider a directory (folder) with such a format:



Here, we will need to crawl whole directory and its sub-directories in order to get a list of all file-paths before writing them to a zip file.
The following program does this by crawling the directory to be zipped:

```
# importing required modules
from zipfile import ZipFile
import os

def get_all_file_paths(directory):

    # initializing empty file paths list
    file_paths = []

    # crawling through directory and subdirectories
    for root, directories, files in os.walk(directory):
```
```

```
        for filename in files:
            # join the two strings in order to form the full filepath.
            filepath = os.path.join(root, filename)
            file_paths.append(filepath)

    # returning all file paths
    return file_paths

def main():
    # path to folder which needs to be zipped
    directory = './python_files'

    # calling function to get all file paths in the directory
    file_paths = get_all_file_paths(directory)

    # printing the list of all files to be zipped
    print('Following files will be zipped:')
    for file_name in file_paths:
        print(file_name)

    # writing files to a zipfile
    with ZipFile('my_python_files.zip','w') as zip:
        # writing each file one by one
        for file in file_paths:
            zip.write(file)

    print('All files zipped successfully!')


if __name__ == "__main__":
    main()
```

The output of above program looks like this:

```
Following files will be zipped:
./python_files\python_basic.pdf
./python_files\python_wiki.txt
./python_files\logos\python_logo.jpg
All files zipped successfully!
```