

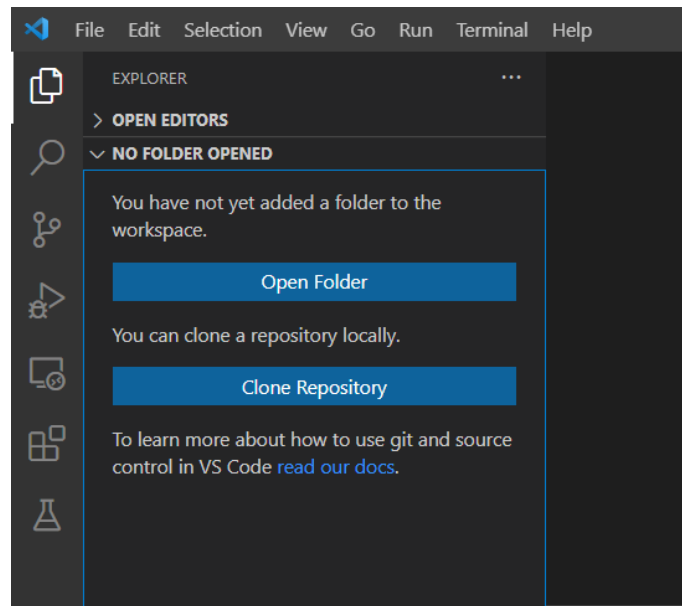
Setup Steps

Please download code and data from: [PyLucene Code and Data](#)

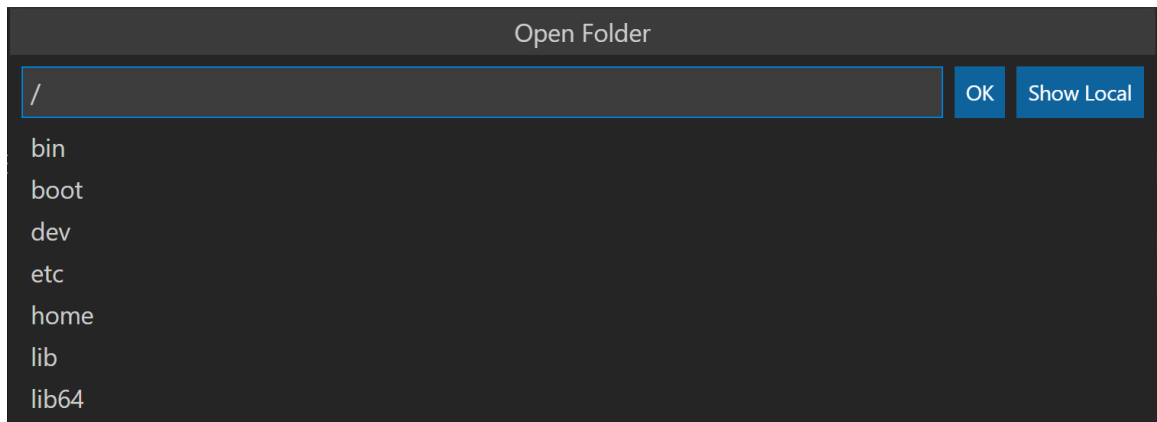
Please check the paths for the commands that have text in blue.

1. Once the code and data has been downloaded, extract the files. We created a folder “Lucene” which contains the folders “Data” and “Code”

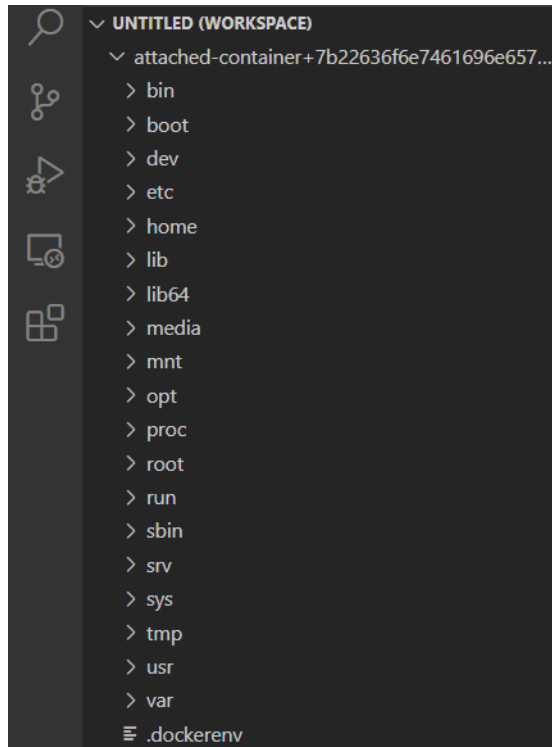
The following are the steps to copy the data to the container. You can go to “Explorer” and click on “Open Folder”



Type “/” and click on “OK”

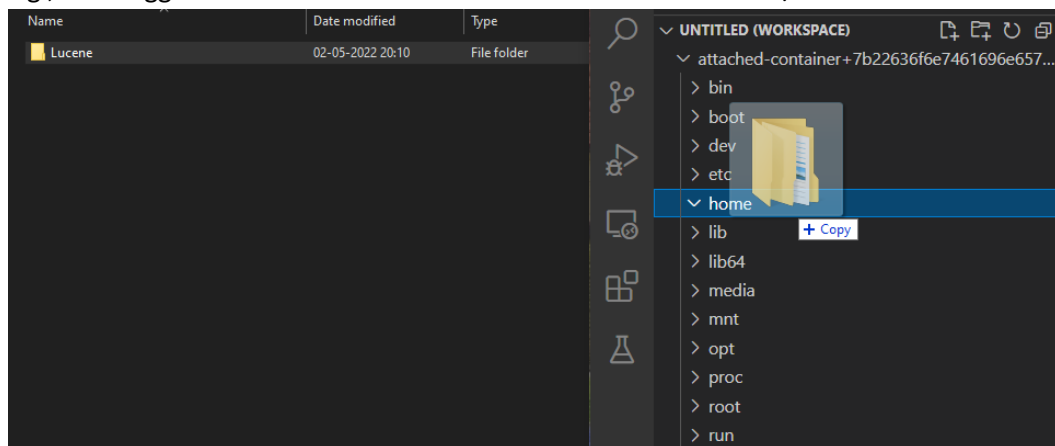


We can see that the workspace has been created under the container.



Now, we can copy the files from our system into the container. The easiest way to do this is to drag and drop the folder into the container.

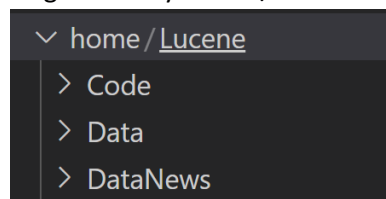
For e.g., we dragged the Lucene folder with the data and code under /home



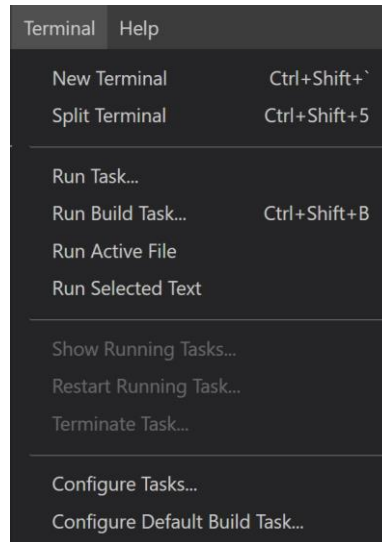
This process can take a while to complete as there are about 42,00 files under DataNews.

If you are unable to see the files, please click on the “refresh explorer” button which appears next to the container name.

You will be able to see the following hierarchy under /home.



2. Then click on Terminal > New Terminal



3. This will open a new blank terminal like the one below:

A screenshot of a terminal window within an IDE. The terminal title bar shows several open files: 'IndexFiles.py', 'SearchFiles.py', 'FieldIndex.py', and 'FieldSearch.py'. The terminal content shows a Python script being executed. The script starts with a shebang line '#!/usr/bin/env python' and defines a constant 'INDEX_DIR = "Index.index"'. It imports several modules: 'sys', 'os', 'lucene', 'threading', 'time', and 'datetime'. It then imports specific classes from the 'lucene' module: 'LimitTokenCountAnalyzer', 'StandardAnalyzer', 'Document', 'Field', 'FieldType', 'FieldInfo', 'IndexWriter', 'IndexWriterConfig', 'IndexOptions', and 'SimpleFSDirectory'. A class 'Ticker' is defined, inheriting from 'object'. It has an '__init__' method that sets 'self.tick = True' and a 'run' method that enters a 'while self.tick:' loop. The terminal output shows the prompt 'root@5f8aad29601d: /# '.