Duplicate File Finder

This Python program searches a directory for duplicate files and moves them to a subdirectory called "Duplicates". It uses the SHA-256 hashing algorithm to compare the contents of the files and detect duplicates. The program also creates a text file called "hashlist.txt" to store the filenames and their corresponding hash values.

Prerequisites

To run this program, you must have Python 3 installed on your system. You can download it from the official website: https://www.python.org/downloads/.

How to Use

1. Download the "duplicate_file_finder.py" file and save it to a directory of your choice.

2. Open a command prompt or terminal window and navigate to the directory where you saved the file.

3. Run the program by entering the following command:

```
Copy code
python duplicate_file_finder.py
```

4. The program will start processing the files in the directory and display a message for each batch of 1000 files. If it finds a duplicate file, it will move it to the "Duplicates" subdirectory and display a message.

5. When the program has finished running, it will create a text file called "hashlist.txt" in the same directory as the program. This file contains a list of the filenames and their corresponding hash values.

Customization

You can customize the behavior of the program by changing the following variables:

- `dir_path`: The directory to search for duplicate files. By default, it is set to the current directory (`os.getcwd()`).

- `duplicates_dir`: The subdirectory to store the duplicate files. By default, it is set to "Duplicates".

- `batch_size`: The number of files to process in each batch. By default, it is set to 1000.

You can also modify the code to use a different hashing algorithm or to exclude certain types of files from the search.

Limitations

This program only detects exact duplicates, i.e., files with identical contents. It does not detect files that are similar but not identical, such as files with different names or metadata. It also does not compare files in subdirectories or across multiple directories.

Acknowledgments

This program was inspired by the "Finding Duplicate Files" project on the Automate the Boring Stuff with Python website: https://automatetheboringstuff.com/chapter9/.