

Generating Data Challenge

Description

This is a test report for document_updater.py. This program processes document files from the originals and updates directories and writes selected files into the finals directory based on the following rules.

Rules

1. If the same file is in both **originals** and **updates** > the one from **updates** should appear in **finals**
 2. If a file in **originals** is on a allowlist, it should appear in **finals**
 3. If a file is on a droplist it shouldn't appear in **finals**
 4. All **updates** always go to **finals** regardless of drop/allow lists
-

Test Cases

All tests were conducted using 4 randomly generated files with names and addresses. To simplify the examples, filenames A, B, C, D are used in this table to represent real files.

ID	Description	Expected behaviour	Actual behaviour
TC01	originals empty, updates has files, droplist/allowlist empty	all update files should appear in finals	As expected
TC02	Both originals and updates empty, droplist/allowlist empty	finals should be empty	As expected

Droplist test cases			
TC03	Files in originals , one file matches droplist, updates empty	File on droplist should not appear in finals All other files should appear in finals	As expected
TC04	All files in originals match droplist, updates empty	Nothing should appear in finals	As expected
TC05	Same name file in both originals and updates , droplist empty	only version from updates should be in finals	As expected

Droplist test cases			
TC06	Same file name in originals , updates and in droplist	updates version should appear in finals	As expected
TC07	originals empty, files in updates , files match droplist	updates files should appear in finals	As expected

Allowlist test cases			
ID	Description	Expected behaviour	Actual behaviour
TC08	Files in originals , one file matches allowlist, updates empty	Only the file on allowlist should appear in finals	As expected
TC09	File in originals matches allowlist, but also present in updates	Only updates version should appear in finals	As expected
TC10	File only in updates , not mentioned in allowlist, originals empty	File should still appear in finals	IndexError: list index out of range (Bug 01)
TC11	Both originals and updates have matching files, allowlist is empty	updates files should appear in finals	IndexError: list index out of range (Bug 01)
TC12	Originals have files, updates are empty, allowlist empty	Finals should be created empty	IndexError: list index out of range (Bug 01)
TC13	Files in originals : ABCD Files in updates : A Allowlist: ABCD	Finals : A (updates versions) BCD(originals)	Only A (updates version) appears in finals (Bug 02)

Special symbols & file names			
ID	Description	Expected behaviour	Actual behaviour
TC14	Hyphenated names - file is in originals ; updates and droplist are empty	File should appear in finals	File does appear in finals but the file name created is missing the hyphen (Bug 03)

Bugs report

The repository used can be cloned from:

<https://github.com/Bara170/generating-data-challenge.git>

Set-up before testing: Install requirements from requirements.txt

Bug 01 - IndexError with allowlist

Description: With allowlist present when there is a file/files in updates folder but allowlist is empty and originals folder is empty, running the program results in an IndexError. Additionally, the finals folder is not created, even though the updates folder contains files that should be copied there.

```
⊗ (data_challenge_venv) → test_data_challenge git:(main) ✗ python document_updater.py data
Traceback (most recent call last):
  File "/Users/barborasrenkova/Documents/test_data_challenge/document_updater.py", line 40, in <module>
    line = list_data[0] if listfiletype == 'allow' else list_data
           ~~~~~~^
```

Expected behaviour: File/ files in updates folder should have been added to finals folder.

Reproduction steps:

1. Navigate to the repository
2. Make sure originals and updates are both empty and there is no finals folder
3. Run "python generate_originals.py"
4. Run "python generate_updates.py"
5. Erase all files from originals folder
6. Ensure that allowlist file exists but is empty
7. Run the program: "python document_updater.py data"

Notes:

- When the file from updates is added to allowlist the program runs fine
- Note that the same error came up for TC11 and TC12 - it seems that this issue is in general caused whenever the allowlist is empty

Bug 02 - Only updates files are copied even though allowlist has additional files from originals

Description: When all files from originals are on allowlist but only one of the files has an updated version in updates - only the updated version is copied over to finals and nothing else

Expected behaviour: All files from originals that are not overridden by updates should also appear in finals

Reproduction steps:

1. Navigate to repository
2. Make sure originals and updates are both empty and there is no finals folder
3. Run "python generate_originals.py"
4. Run "python generate_updates.py"
5. Make sure allowlist exists and copy all file names from originals into it
6. Only leave 1 file in updates and remove the rest
7. Run the program: "python document_updater.py data"

Bug 03 - Hyphenated names not used correctly in the file names

Description: When a hyphenated surname is present, the file that is copied into finals doesn't include the hyphen in the file name

Expected Behaviour: The name of the file created in finals should include the hyphen

Reproduction steps:

1. Navigate to the repository
2. Make sure originals and updates are both empty and there is no finals folder
3. Run "python generate_originals.py"
4. Run "python generate_updates.py"
5. Ensure that droplist(if allowlist present, rename it to droplist) exists but is empty
6. In the updates folder - change the file name "Hayes" to "Hayes-Smith"
7. Open the file and change the person's surname to "Hayes-Smith"
8. Run the program: "python document_updater.py data"
9. The file name created in finals will be showing as HayesSmith