git-filter-repo

Install git-filter-repo

Download file and save it without extension "git-filter-repo":

https://github.com/newren/git-filter-repo/blob/main/git-filter-repo

Instructions to install "git-filter-repo:

https://github.com/newren/git-filter-repo/tree/main?tab=readme-ov-file#how-do-i-install-it

Instructions for windows users "git-filter-repo":

https://github.com/newren/git-filter-repo/blob/main/INSTALL.md#notes-for-windows-users

Start using git-filter-repo

Note: git-filter-repo needs python.

In my case using GitBash (in Windows laptop):

- First, clone the repository to filter → **Always keep a copy of the original repository!!!**
- File "git-filter-repo" must be in the cloned repository.
- Run commands git-filter-repo. Here you can find the man pages:
https://htmlpreview.github.io/?https://github.com/newren/git-filter-repo/blob/docs/html/git-filter-repo.html
https://www.mankier.com/1/git-filter-repo

Example of a command:

git filter-repo --path path/to/folder_to_filter

In GitBash (windows laptop):

python git-filter-repo --path path/to/folder_to_filter

Tutorial

Original repository: ToyRepo, with 10 subfolders

Filtered repository: ToyRepo_filter_00, with only subfolder "9" and some files.

List the commit history of ToyRepo (to compare later with the filtered repo):

```
Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo (main)
$ git log
commit 5c60b38ec8e952995b6dbb87a636b0c3f5f3a6be (HEAD -> main, origin/main)
Author: AGonzalezNicolas <a.gonzalez-nicolas@fz-juelich.de>
Date:   Tue Jun 11 14:39:34 2024 +0200

    Modify file in folder_7 again to try update filtered_03 without cloning first.

commit 20daba166c918c88cc07e4da43570888ed488ada
Author: AGonzalezNicolas <a.gonzalez-nicolas@fz-juelich.de>
Date:   Tue Jun 11 13:57:07 2024 +0200

    Modify files inside (_7) & outside (_0) filter

commit db60625f85d97d85db8fd6859a15ec727b29086c
Author: AGonzalezNicolas <a.gonzalez-nicolas@fz-juelich.de>
Date:   Tue Jun 11 13:06:09 2024 +0200

    Add git-filter-repo file in Original ToyRepo

commit be30a0ba28f62179e18e93cc46d0c35775d97885
Author: AGonzalezNicolas <a.gonzalez-nicolas@fz-juelich.de>
Date:   Tue Jun 11 10:23:31 2024 +0200

    Modify one file inside filter: Folder_5/samples_5.dat

commit 335dc2cb93e3bc913adcadda79c3afed453744ca
Author: AGonzalezNicolas <a.gonzalez-nicolas@fz-juelich.de>
Date:   Tue Jun 11 10:21:32 2024 +0200

    Modify one file outside filter: Folder_1/samples_1.dat

commit b074d4f44f7885192b1e704cf4cd005a90a86c6f
Author: AGonzalezNicolas <a.gonzalez-nicolas@fz-juelich.de>
Date:   Tue Jun 11 08:49:28 2024 +0200

    Changes in samples Folder_5

commit aebc590e189d34c9015fba71cea102431998ec4f
Author: AGonzalezNicolas <a.gonzalez-nicolas@fz-juelich.de>
Date:   Mon Jun 10 14:58:38 2024 +0200

    Commit files not in subfolders

commit d3806ff9288c41b4c084549d97f65caef3a6f47e
Author: AGonzalezNicolas <a.gonzalez-nicolas@fz-juelich.de>
Date:   Mon Jun 10 14:57:17 2024 +0200

    Commit folders from 3 to 9

commit 526a8d95c6f905486c1680aa86116ccec44f9c9b
Author: AGonzalezNicolas <a.gonzalez-nicolas@fz-juelich.de>
Date:   Mon Jun 10 14:56:29 2024 +0200

    Commit folders 0, 1, 2
```

Filter one repository

- Clone locally original repo ToyRepo:

git clone git@github.com:AGonzalezNicolas/ToyRepo.git ToyRepo_toFilter_00

- Move to the cloned repo directory and list content of this repo before filtering:
  cd ToyRepo_toFilter_00

```
Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo_toFilter_00 (main)
$ ls -lsa
total 202
    4 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 ./
    4 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:27 ../
    4 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 .git/
    0 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 0/
    0 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 1/
    0 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 2/
    0 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 3/
    0 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 4/
    0 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 5/
    0 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 6/
    0 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 7/
    0 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 8/
    0 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 9/
    1 -rwxr-xr-x 1 Ana 197121    353 Jun 12 09:28 create_foldersfiles*
   20 -rw-r--r-- 1 Ana 197121  16460 Jun 12 09:28 default_single.tcl
  168 -rwxr-xr-x 1 Ana 197121 169787 Jun 12 09:28 git-filter-repo*
    1 -rw-r--r-- 1 Ana 197121    500 Jun 12 09:28 samples.dat
```

- Filter the folder of interest ("9") and some files (default_single.tcl, samples.dat, git-filter-repo):
  python git-filter-repo --path default_single.tcl --path samples.dat --path git-filter-repo --path 9/

- List again content of this filtered repo. Now the filtered repository only includes the files and folder we wanted:

```
Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo_toFilter_00 (main)
$ ls -lsa
total 201
    4 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:32 ./
    4 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:27 ../
    4 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:32 .git/
    0 drwxr-xr-x 1 Ana 197121      0 Jun 12 09:28 9/
   20 -rw-r--r-- 1 Ana 197121  16460 Jun 12 09:28 default_single.tcl
  168 -rwxr-xr-x 1 Ana 197121 169787 Jun 12 09:28 git-filter-repo*
    1 -rw-r--r-- 1 Ana 197121    500 Jun 12 09:28 samples.dat
```

- The commits on the commit history of this filtered repo are only related to the files and folder_9:

```
Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo_toFilter_00 (main)
$ git log
commit be3068248c466d0e00c9dfdbfcb0d6584332e2fe (HEAD -> main)
Author: AGonzalezNicolas <a.gonzalez-nicolas@fz-juelich.de>
Date:   Tue Jun 11 13:06:09 2024 +0200

    Add git-filter-repo file in Original ToyRepo

commit e1019aface4323ad2884a315fbcb81b9b570376c
Author: AGonzalezNicolas <a.gonzalez-nicolas@fz-juelich.de>
Date:   Mon Jun 10 14:58:38 2024 +0200

    Commit files not in subfolders

commit eaf45dffb7b2f38c7f86d5644f3ddf19052f2a1d
Author: AGonzalezNicolas <a.gonzalez-nicolas@fz-juelich.de>
Date:   Mon Jun 10 14:57:17 2024 +0200

    Commit folders from 3 to 9
```

- If you check the remotes of this filtered repo, you will see there are not origins/remote repositories associated:

```
Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo_toFilter_00 (main)
$ git remote -v

Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo_toFilter_00 (main)
$
```

Changes on original repository

Move to the folder where the original repository `ToyRepo` is.

- Make some changes on the original repository. In my case I will change locally:
    o Outside filter: file `1/samples_1.dat`
    o Inside filter: file `9/samples_9.dat`

- Add changes and commit them. Commit them separately, so you have two different commits and
  you will "take" afterwards only the commit related to the filter (folder_9) and not the commit
  outsider the filter (folder_1).
  git add 1/samples_1.dat
  git commit -m "Changes in 1/samples_1.dat on 12.06.2024"
  git add 9/samples_9.dat
  git commit -m "Changes in 9/samples_9.dat on 12.06.2024"

- Check commits on original repo:

```
Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo (main)
$ git log --oneline
5605948 (HEAD -> main) Changes in 9/samples_9.dat on 12.06.2024
529b281 Changes in 1/samples_1.dat on 12.06.2024
5c60b38 (origin/main) Modify file in folder_7 again to try update filtered_03 without cloning first.
20daba1 Modify files inside (_7) & outside (_0) filter
db60625 Add git-filter-repo file in Original ToyRepo
be30a0b Modify one file inside filter: Folder_5/samples_5.dat
335dc2c Modify one file outside filter: Folder_1/samples_1.dat
b074d4f Changes in samples Folder_5
aebc590 Commit files not in subfolders
d3806ff Commit folders from 3 to 9
526a8d9 Commit folders 0, 1, 2
```

- Then push changes to the remote repository:

  git push origin main

You can also check that those changes appear on your remote original repository (ToyRepo):

Changes on filtered repository

Move to the filtered repository `ToyRepo_toFilter_00` folder.

- Modify the same file (as we did in the original repo `ToyRepo`) in the filtered repository: `9/sample_9.dat`

- Add and commit this change:

  git add 9/samples_9.dat

  git commit -m "Filter repository: Changes in 9/samples_9.dat on 12.06.2024"

- Check commit history:

  git log --oneline

Merge original and filtered repository

Continue working on the filtered repository locally.

- Check remote origins on the filtered repo (there shouldn't be any):

```
Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo_toFilter_00 (main)
$ git remote -v

Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo_toFilter_00 (main)
$
```

- Add as remote the original repository ToyRepo:

  git remote add ToyRepo git@github.com:AGonzalezNicolas/ToyRepo.git

```
Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo_toFilter_00 (main)
$ git remote -v
ToyRepo git@github.com:AGonzalezNicolas/ToyRepo.git (fetch)
ToyRepo git@github.com:AGonzalezNicolas/ToyRepo.git (push)
```

- Fetch changes from the original repository:

  git fetch ToyRepo main

- You can check branches:

  git branch -a

```
Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo_toFilter_00 (main)
$ git branch -a
* main
  remotes/ToyRepo/main
```

- Merge what you fetched:

  git merge ToyRepo/main main

```
Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo_toFilter_00 (main)
$ git merge ToyRepo/main main
Auto-merging 9/samples_9.dat
CONFLICT (content): Merge conflict in 9/samples_9.dat
Automatic merge failed; fix conflicts and then commit the result.

Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo_toFilter_00 (main|MERGING)
$
```

- Solve conflicts on file `Samples_9.dat` since it was modified on both repositories (original and filtered repos). I used Visual Studio Code to solve the conflict (accepted both changes).

- Then commit:

  git commit -m "Add changes from original repository after development on the filtered repo"

  - In the log you should be able to see the commit related to folder_9 in the original repo (but not related to folder_1).

```
Ana@ZAM445 MINGW64 /c/0_Ana_Julich/GIT_repos/0_ToyRepo/ToyRepo_toFilter_00 (main)
$ git log --oneline
9325a7e (HEAD -> main) Add changes from original repository after development on the filtered repo
33a17ce Filter repository: Changes in 9/samples_9.dat on 12.06.2024
5605948 (replaced, ToyRepo/main) Changes in 9/samples_9.dat on 12.06.2024
a066d4c Modify files inside (_7) & outside (_0) filter
e1019af Commit files not in subfolders
eaf45df Commit folders from 3 to 9
```

Note: after the merge, you fetch all folders and files from the original repo. Just apply the same filter as in the beginning to filter files/folders and commit history.

python git-filter-repo --path default_single.tcl --path samples.dat --path git-filter-repo --path 9/

More info:

> https://stackoverflow.com/questions/69632636/updating-git-repo-after-using-git-filter-repo-subdirectory-filter