How do I split a file into two while preserving git line history?

devblogs.microsoft.com/oldnewthing/20190916-00

September 16, 2019

Raymond Chen

Some time ago, I showed how to combine two files into one while preserving line history. Today, we're going to do the opposite: Split a file into two smaller files, while preserving line history.

Let's set up a scratch repo to demonstrate. I've omitted the command prompts so you can copy-paste this into your shell of choice and play along at home. (The timestamps and commit hashes will naturally be different.)

```
git init

>foods echo apple
>>foods echo celery
>>foods echo cheese
git add foods
git commit --author="Alice <alice>" -m created

>>foods echo eggs
>>foods echo grape
>>foods echo lettuce
git commit --author="Bob <bob>"   -am middle

>>foods echo milk
>>foods echo orange
>>foods echo peas
git commit --author="Carol <carol>" -am last

git tag ready
```

With this starting point, the `git blame` output says

```
^e7a114d (Alice 2019-09-16 07:00:00 -0700 1) apple
^e7a114d (Alice 2019-09-16 07:00:00 -0700 2) celery
^e7a114d (Alice 2019-09-16 07:00:00 -0700 3) cheese
86348be4 (Bob   2019-09-16 07:00:01 -0700 4) eggs
86348be4 (Bob   2019-09-16 07:00:01 -0700 5) grape
86348be4 (Bob   2019-09-16 07:00:01 -0700 6) lettuce
34eb5bd1 (Carol 2019-09-16 07:00:02 -0700 7) milk
34eb5bd1 (Carol 2019-09-16 07:00:02 -0700 8) orange
34eb5bd1 (Carol 2019-09-16 07:00:02 -0700 9) peas
```

As we noted when we learned how to combine two files, the naïve way of splitting the file will treat the larger file as a continuation of the original (assuming you haven't hit the rename limit), and the smaller file will be treated as a brand new file. The blame of the smaller file will blame you, the person who split them, instead of blaming the person who introduced each line.

To get git to follow the line attributes, we have to make each of the result files look like a rename of the original. We can do this by creating each piece in a different branch, then merging them all together.

In a new `fruits` branch, the first step is to do a pure rename, so that git will recognize that the `fruits` file is a continuation of the `foods` file.

```
git checkout -b fruits
git mv foods fruits
git commit --author="Greg <greg>" -m "split foods to fruits"
```

Now you can edit the `fruits` file to contain just the part you want to split out. In this case, we want the fruits (duh).

```
>fruits echo apple
>>fruits echo grape
>>fruits echo orange
git commit --author="Greg <greg>" -am "split foods to fruits"

git checkout -
```

Repeat for the veggies.

```
git checkout -b veggies
git mv foods veggies
git commit --author="Greg <greg>" -m "split foods to veggies"

>veggies echo celery
>>veggies echo lettuce
>>veggies echo peas
git commit --author="Greg <greg>" -am "split foods to veggies"

git checkout -
```

The last file (dairy) can be done directly in the original branch.

```
git mv foods dairy
git commit --author="Greg <greg>" -m "split foods to dairy"

>dairy echo cheese
>>dairy echo eggs
>>dairy echo milk
git commit --author="Greg <greg>" -am "split foods to dairy"
```

And now we octopus merge all the branches together.

```
git merge fruits veggies
```

This time, the octopus merge succeeds. All branches agree that the `foods` file be deleted, so there are no merge conflicts.

```
Trying simple merge with fruits
Trying simple merge with veggies
Merge made by the 'octopus' strategy.
 fruits  | 3 +++
 veggies | 3 +++
 2 files changed, 6 insertions(+)
 create mode 100644 fruits
 create mode 100644 veggies
```

And lo and behold, all three resulting files preserved the original line histories. Greg doesn't show up anywhere.

```
git blame fruits

^e7a114d foods (Alice 2019-09-16 07:00:00 -0700 1) apple
86348be4 foods (Bob   2019-09-16 07:00:01 -0700 2) grape
34eb5bd1 foods (Carol 2019-09-16 07:00:02 -0700 3) orange

git blame veggies

^e7a114d foods (Alice 2019-09-16 07:00:00 -0700 1) celery
86348be4 foods (Bob   2019-09-16 07:00:01 -0700 2) lettuce
34eb5bd1 foods (Carol 2019-09-16 07:00:02 -0700 3) peas

git blame dairy

^e7a114d foods (Alice 2019-09-16 07:00:00 -0700 1) cheese
86348be4 foods (Bob   2019-09-16 07:00:01 -0700 2) eggs
34eb5bd1 foods (Carol 2019-09-16 07:00:02 -0700 3) milk
```

Raymond Chen

Follow