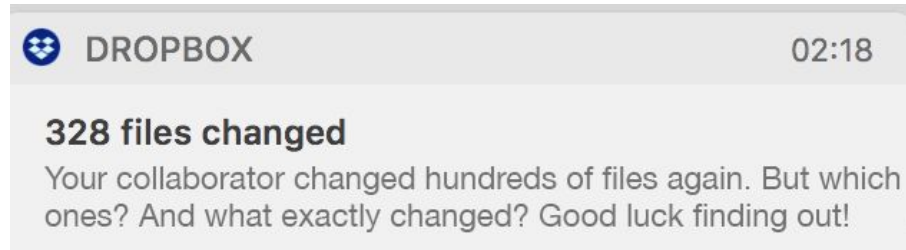


Versioning Using Git

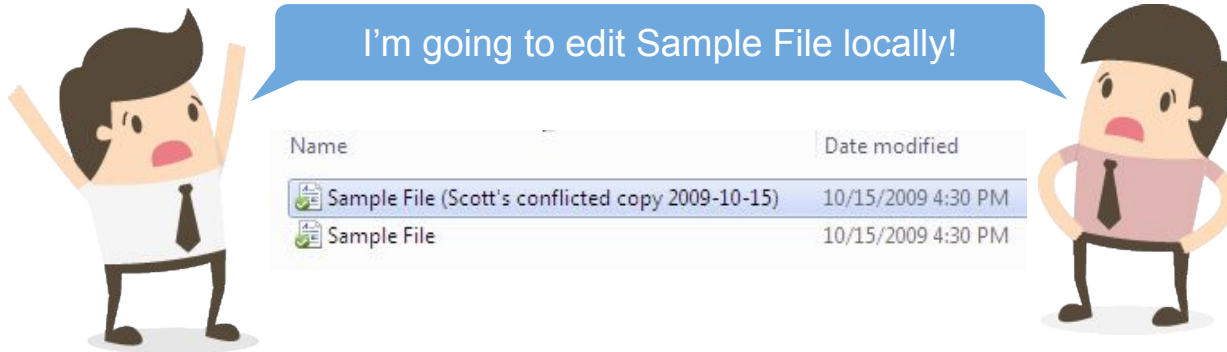
By **Pieter Delobelle** & Thomas Winters

Working with code

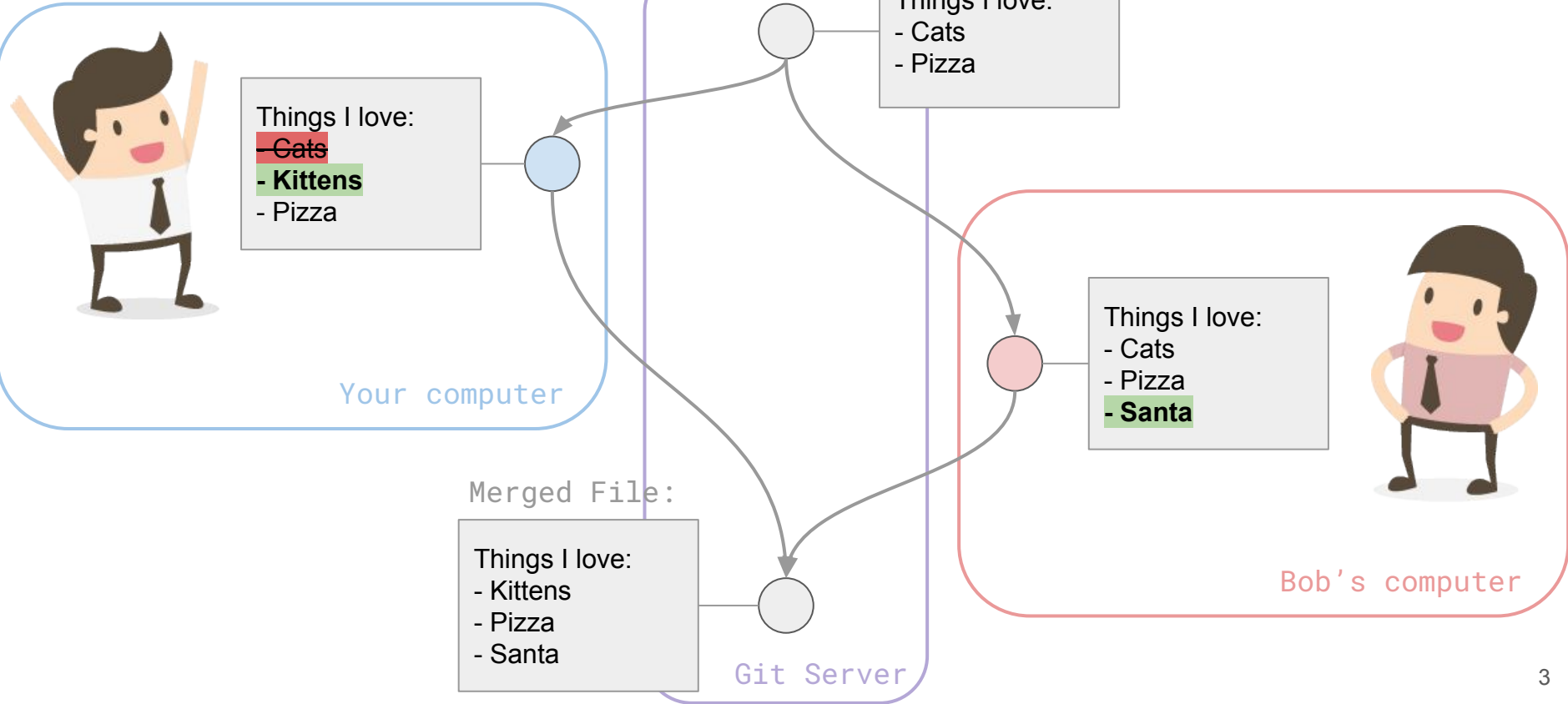
One might use classic cloud services for code sharing & backups (e.g. Dropbox),
BUT:



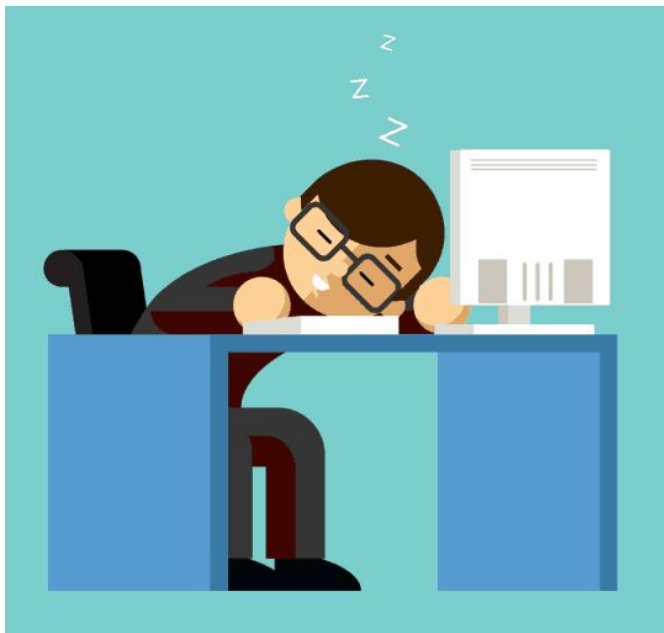
Or worse: simultaneous edits leading to conflicted copies



Git to the rescue



But still useful when working on your own



Not-so-hypothetical situation

- Updating your scripts one late evening
- Next day, something broke

Questions:

- What broke where? Reset all progress from last night to get working version again?
- Which lines changed exactly?

→ Git to the rescue!



How to share code & models afterwards?

Dropbox-like cloud storage?

BUT: What if full, or after quitting academia?

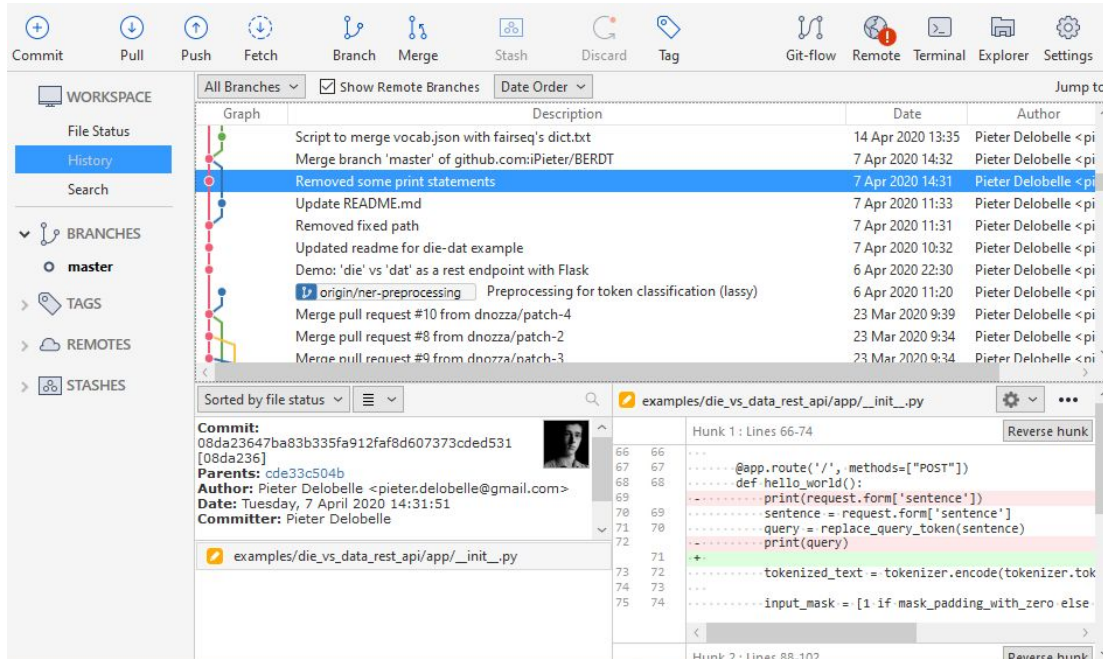
Upload to own website?

BUT: Lot of work, and less discoverability

What if want to discuss with downloaders & easily release new versions?

Create a [Git repository](#)!

Getting visual overview on your repository: SourceTree



The screenshot displays the SourceTree application interface. At the top, there is a toolbar with icons for Commit, Pull, Push, Fetch, Branch, Merge, Stash, Discard, Tag, Git-flow, Remote, Terminal, Explorer, and Settings. Below the toolbar, the main window is divided into several sections:

- Workspace:** Includes File Status, History (highlighted), and Search.
- BRANCHES:** Shows the current branch as 'master'.
- TAGS:** A section for tags.
- REOTES:** A section for remote repositories.
- STASHES:** A section for stashed changes.

The central area shows a commit history table with columns for Graph, Description, Date, and Author. The selected commit is:

Graph	Description	Date	Author
[Red circle]	Removed some print statements	7 Apr 2020 14:31	Pieter Delobelle <pi

Below the commit history, the file status is shown as 'examples/die_vs_data_rest_api/app/_init_.py'. The commit details are:

Commit: 08da23647ba83b335fa912faf8d607373cded531 [08da236]
Parents: cde33c504b
Author: Pieter Delobelle <pieter.delobelle@gmail.com>
Date: Tuesday, 7 April 2020 14:31:51
Committer: Pieter Delobelle

The diff view shows the changes in the file, with a hunk of code highlighted in red and green:

```
Hunk 1: Lines 66-74
Reverse hunk
66 66 ...
67 67 ..... @app.route('/', methods=["POST"])
68 68 ..... def hello_world():
69 69 .....     print(request.form['sentence'])
70 69 .....     sentence = request.form['sentence']
71 70 .....     query = replace_query_token(sentence)
72 70 .....     print(query)
73 72 .....     tokenized_text = tokenizer.encode(tokenizer.tok
74 73 .....
75 74 .....     input_mask = [1 if mask_padding_with_zero else
Reverse hunk
Hunk 2: Lines 88-102
Reverse hunk
```

Use a nice GUI (e.g. SourceTree) for getting an easy overview to your repository

<https://www.sourcetreeapp.com/>

(Command line is also often used, but less accessible for beginners)

Git Terminology

The layout of a Git Project

Repository

Your code for one project lives in a **repository**

Repository: my-super-awesome-project

Committing your changes

Every time you “save” your code, you create a new **commit**

Repository: my-super-awesome-project



Commit 1

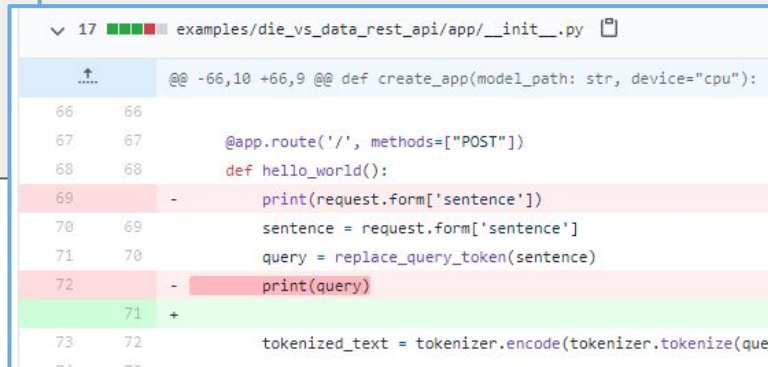
Initial commit

Committing your changes

Every time you “save” your code, you create a new **commit**

Repository: my-super-awesome-project

Commit 1
Initial commit



```
17 examples/die_vs_data_rest_api/app/__init__.py
↑  @@ -66,10 +66,9 @@ def create_app(model_path: str, device="cpu"):
66 66
67 67     @app.route('/', methods=["POST"])
68 68     def hello_world():
69 - print(request.form['sentence'])
70 69     sentence = request.form['sentence']
71 70     query = replace_query_token(sentence)
72 - print(query)
71 +
73 72     tokenized_text = tokenizer.encode(tokenizer.tokenize(que
```

Code branches

A lineage of commits is called a **branch**, which refers to the latest commit
For example the **main branch**



Code branches

You have a **main branch** (or **master**) and you can add other named branches for creating variations of your code

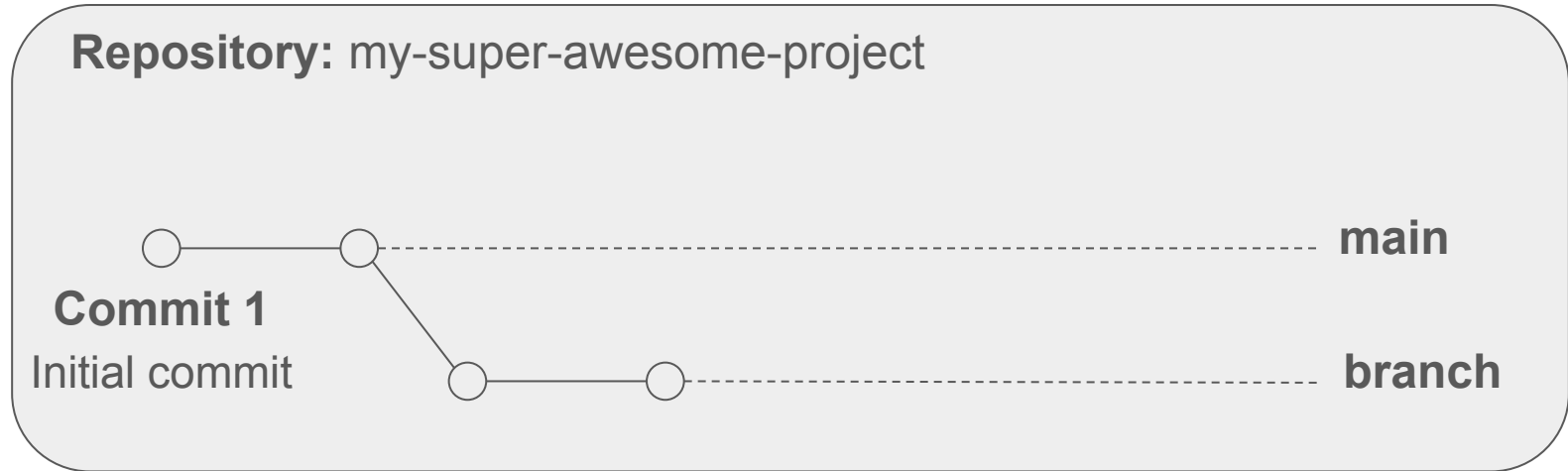


Code branches

You can commit on this new branch without affecting the main branch other collaborators might simultaneously be working on

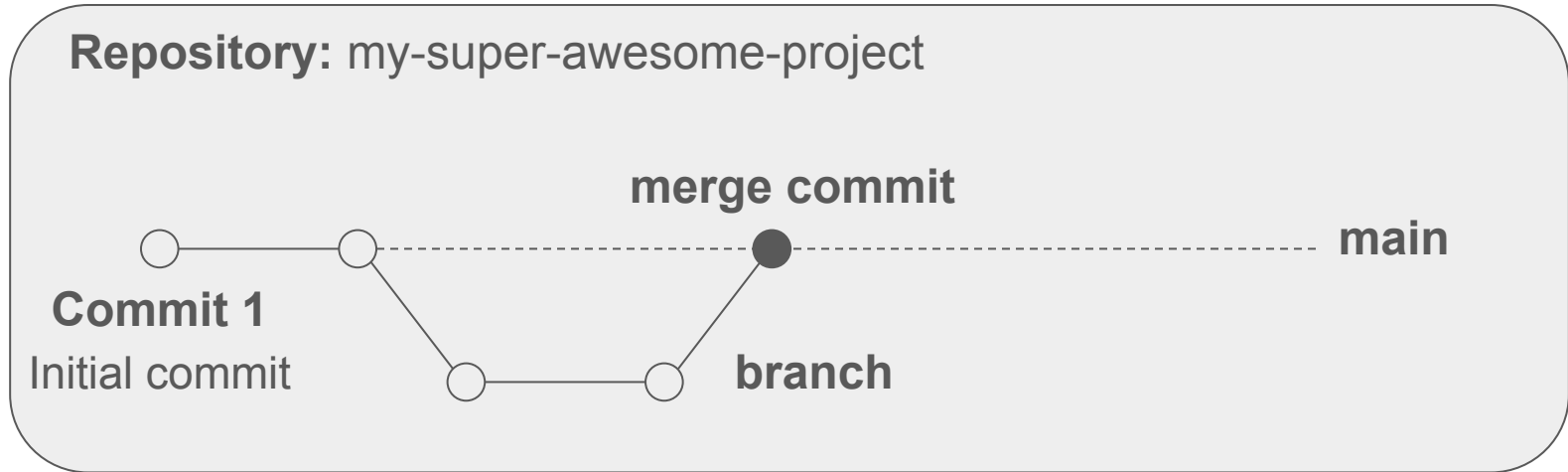


Code branches

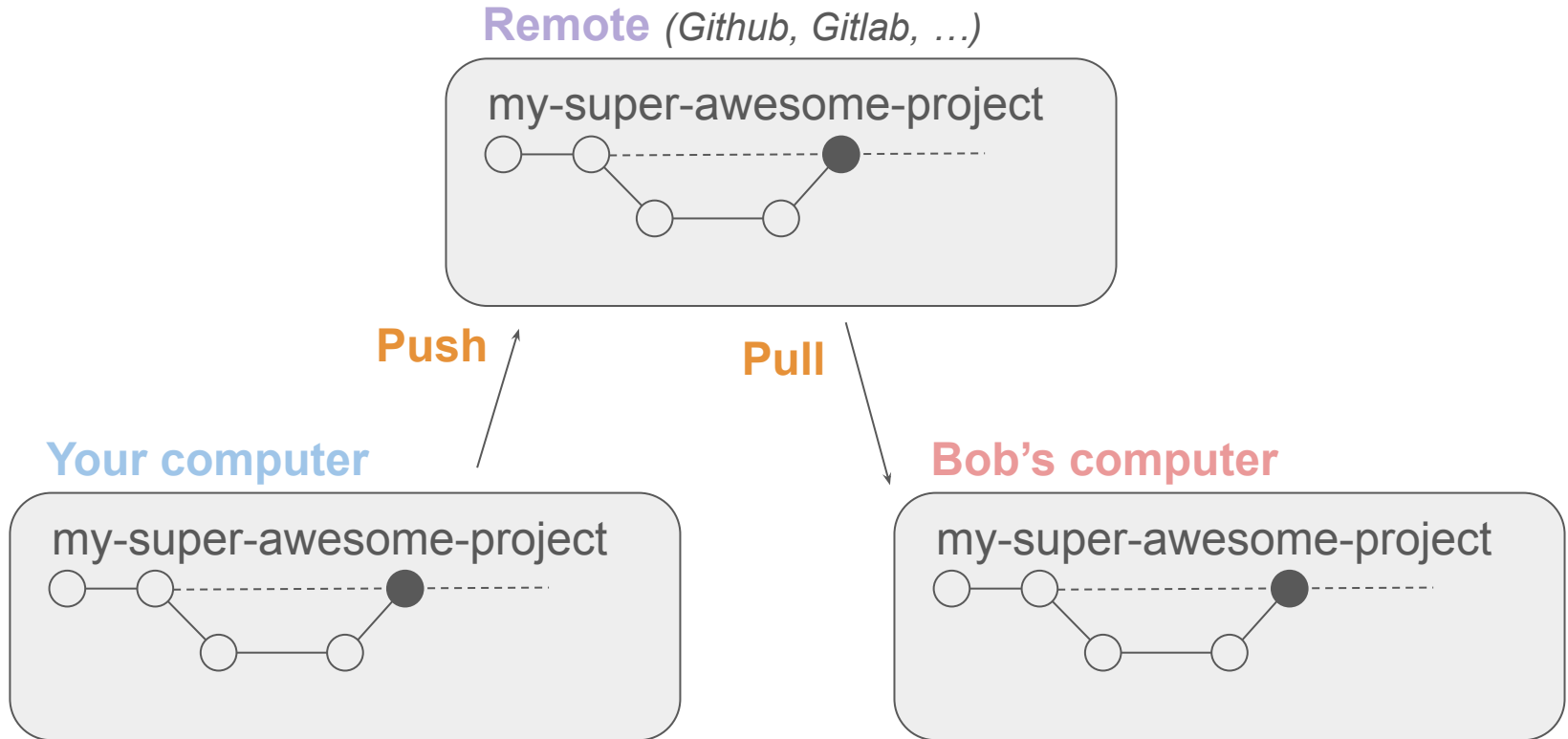


Merging branches

And when the code is ready, you can put it back on master with a **merge commit**



Push your changes to remote server, and pull for latest



GUI provides easy access to all these concepts

The screenshot displays the Git GUI interface. At the top, a toolbar contains icons for Commit, Pull, Push, Fetch, Branch, Merge, Stash, Discard, Tag, Git-flow, Remote, Terminal, Explorer, and Settings. The Commit, Pull, Push, Fetch, Branch, and Merge icons are highlighted with a red box. Below the toolbar, the left sidebar shows the workspace structure with sections for File Status, History, Search, BRANCHES (master), TAGS, REMOTES, and STASHES. The main area is divided into two panes. The upper pane shows a commit history table with columns for Graph, Description, Date, and Author. The lower pane shows a code diff for the file 'examples/die_vs_data_rest_api/app/__init__.py', displaying line numbers and code changes.

Graph	Description	Date	Author
	Script to merge vocab.json with fairseq's dict.txt	14 Apr 2020 13:35	Pieter Delobelle <pi
	Merge branch 'master' of github.com:iPieter/BERDT	7 Apr 2020 14:32	Pieter Delobelle <pi
	Removed some print statements	7 Apr 2020 14:31	Pieter Delobelle <pi
	Update README.md	7 Apr 2020 11:33	Pieter Delobelle <pi
	Removed fixed path	7 Apr 2020 11:31	Pieter Delobelle <pi
	Updated readme for die-dat example	7 Apr 2020 10:32	Pieter Delobelle <pi
	Demo: 'die' vs 'dat' as a rest endpoint with Flask	6 Apr 2020 22:30	Pieter Delobelle <pi
	origin/ner-preprocessing Preprocessing for token classification (lassy)	6 Apr 2020 11:20	Pieter Delobelle <pi
	Merge pull request #10 from dnozza/patch-4	23 Mar 2020 9:39	Pieter Delobelle <pi
	Merge pull request #8 from dnozza/patch-2	23 Mar 2020 9:34	Pieter Delobelle <pi
	Merge pull request #9 from dnozza/patch-3	23 Mar 2020 9:34	Pieter Delobelle <pi

Sorted by file status

examples/die_vs_data_rest_api/app/__init__.py

Commit: 08da23647ba83b335fa912faf8d607373cded531 [08da236]
Parents: cde33c504b
Author: Pieter Delobelle <pieter.delobelle@gmail.com>
Date: Tuesday, 7 April 2020 14:31:51
Committer: Pieter Delobelle

Hunk 1: Lines 66-74

```
66 66 ...  
67 67 ..... @app.route('/', methods=["POST"])  
68 68 ..... def hello_world():  
69 69 .....     print(request.form['sentence'])  
70 69 .....     sentence = request.form['sentence']  
71 70 .....     query = replace_query_token(sentence)  
72 70 .....     print(query)  
73 72 .....     tokenized_text = tokenizer.encode(tokenizer.tok  
74 73 .....  
75 74 .....     input_mask = [1 if mask_padding_with_zero else
```

Who can host your repositories?

KU Leuven servers: gitlab.kuleuven.be

Third party service: github.com, gitlab.com, ... *(if project NDA allows)*

Interface

sourcetreeapp.com

Further tutorials

try.github.io/