

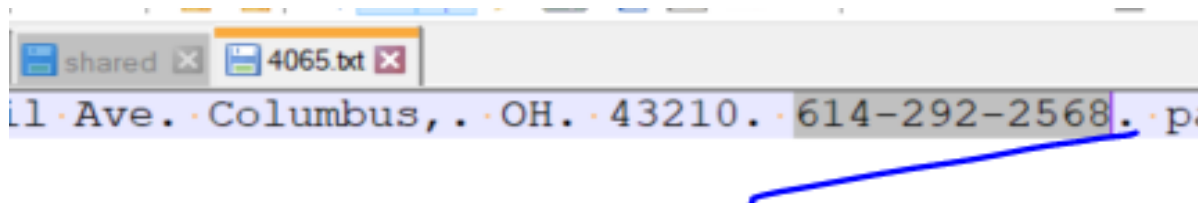
Step 1: Select a txt to test



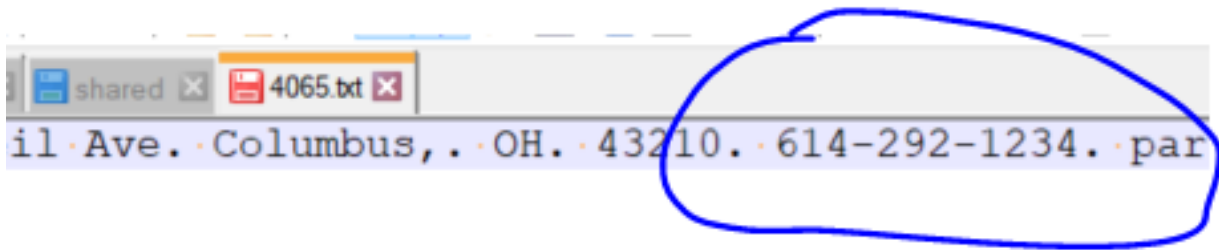
Step 2: Find the txt file id

```
--end----- line_58
*****begin***** line_96
('results size:', 5)
('results contents:', '[(4065L, 0.6512059569358826), (6086L, 0.386
.3659544587135315), (4078L, 0.36383184790611267)]')
('results[0]', (4065L, 0.6512059569358826))
('results[1]', (6086L, 0.3868231177330017))
('results[2]', (6111L, 0.3797479271888733))
('results[3]', (6100L, 0.3659544587135315))
('results[4]', (4078L, 0.36383184790611267))
*****end***** line_96
+++++
```

Step 3: Open the txt file



Step 4: Change the phone number



Step 5: Preprocess the txt files again. You need to specify the max txt file id as 6524

is PC > Windows (C:) > peter > code > git_local > data > compiled_bios

Name	Date modified	Type	Size
6490.txt	2021-12-02 6:52 PM	Text Document	4 KB
6491.txt	2021-12-02 6:52 PM	Text Document	4 KB
6492.txt	2021-12-02 6:52 PM	Text Document	9 KB
6493.txt	2021-12-02 6:52 PM	Text Document	2 KB
6494.txt	2021-12-02 6:52 PM	Text Document	19 KB
6495.txt	2021-12-02 6:52 PM	Text Document	4 KB
6496.txt	2021-12-02 6:52 PM	Text Document	23 KB
6497.txt	2021-12-02 6:52 PM	Text Document	3 KB
6498.txt	2021-12-02 6:52 PM	Text Document	4 KB
6499.txt	2021-12-02 6:52 PM	Text Document	1 KB
6500.txt	2021-12-02 6:52 PM	Text Document	6 KB
6501.txt	2021-12-02 6:52 PM	Text Document	4 KB
6502.txt	2021-12-02 6:52 PM	Text Document	1 KB
6503.txt	2021-12-02 6:52 PM	Text Document	5 KB
6504.txt	2021-12-02 6:52 PM	Text Document	5 KB
6505.txt	2021-12-02 6:52 PM	Text Document	5 KB
6506.txt	2021-12-02 6:52 PM	Text Document	4 KB
6507.txt	2021-12-02 6:52 PM	Text Document	2 KB
6508.txt	2021-12-02 6:52 PM	Text Document	95 KB
6509.txt	2021-12-02 6:52 PM	Text Document	3 KB
6510.txt	2021-12-02 6:52 PM	Text Document	6 KB
6511.txt	2021-12-02 6:52 PM	Text Document	2 KB
6512.txt	2021-12-02 6:52 PM	Text Document	8 KB
6513.txt	2021-12-02 6:52 PM	Text Document	6 KB
6514.txt	2021-12-02 6:52 PM	Text Document	3 KB
6515.txt	2021-12-02 6:52 PM	Text Document	3 KB
6516.txt	2021-12-02 6:52 PM	Text Document	25 KB
6517.txt	2021-12-02 6:52 PM	Text Document	23 KB
6518.txt	2021-12-02 6:52 PM	Text Document	2 KB
6519.txt	2021-12-02 6:52 PM	Text Document	13 KB
6520.txt	2021-12-02 6:52 PM	Text Document	2 KB
6521.txt	2021-12-02 6:52 PM	Text Document	2 KB
6522.txt	2021-12-02 6:52 PM	Text Document	3 KB
6523.txt	2021-12-02 6:52 PM	Text Document	5 KB
6524.txt	2021-12-02 6:52 PM	Text Document	3 KB
config.toml	2021-12-02 8:24 PM	TOML File	1 KB
dataset-full_corpus.txt	2021-12-02 8:24 PM	Text Document	101 KB

Step 6: Run the preprocess script `./preprocess.sh 6524`

```

85 history
peter@LAPTOP-MSKOE3G5:/mnt/c/peter/code/git_local$ pwd
/mnt/c/peter/code/git_local
peter@LAPTOP-MSKOE3G5:/mnt/c/peter/code/git_local$ ./preprocess.sh 6524
Max txt file id: 6524
***** Begin to preprocess data.*****
/mnt/c/peter/code/git_local
+++++ Begin to extract information.+++++
Max txt file id: 6524
/mnt/c/peter/code/git_local

Begin to process emails, the source is at: /mnt/c/peter/code/git_local/data/compiled_bios
The destination is at : /mnt/c/peter/code/git_local/data/emails
File: /mnt/c/peter/code/git_local/data/compiled_bios/0.txt

```

Step 7: Start web service

```
gunicorn server:app -b 127.0.0.1:8095
```

```

peter@LAPTOP-MSKOE3G5:/mnt/c/peter/code/git_local$ 
peter@LAPTOP-MSKOE3G5:/mnt/c/peter/code/git_local$ gunicorn server:app -b 127.0.0.1:8095
[2021-12-02 20:28:11 +0000] [156] [INFO] Starting gunicorn 19.7.1
[2021-12-02 20:28:11 +0000] [156] [INFO] Listening at: http://127.0.0.1:8095 (156)
[2021-12-02 20:28:11 +0000] [156] [INFO] Using worker: sync
[2021-12-02 20:28:11 +0000] [160] [INFO] Booting worker with pid: 160

```

Step 8: You can find the number has been changed.

127.0.0.1:8095

rch

data mining

Srinivasan Parthasarathy Phone:614-292-1234 ✓

Computer Science, Ohio State University Ohio, United States

...to steering committee of siam **data mining**. january, 2010. lumley..of ieee international conference on **data mining** 2006.. january, 2007....january, 2005. best of siam **data mining**. january, 2004. nsf

Computer Science, University of Kentucky Kentucky, United States

: research interests: **data mining** machine learning artificial