

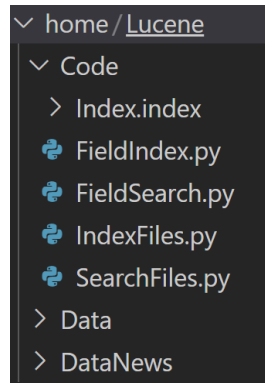
Project Run Steps

1. In the terminal, type: `python /home/Lucene/Code/IndexFiles.py /home/Lucene/Data`

```
root@5e62a4ed2c7d:/# python /home/Lucene/Code/IndexFiles.py /home/Lucene/Data
lucene 8.11.0
adding talk.politics.mideast.txt
adding sci.space.txt
adding sci.electronics.txt
adding sci.med.txt
adding comp.sys.ibm.pc.hardware.txt
adding comp.graphics.txt
adding talk.politics.guns.txt
adding comp.windows.x.txt
adding rec.autos.txt
adding soc.religion.christian.txt
adding rec.sport.hockey.txt
adding talk.religion.misc.txt
adding rec.sport.baseball.txt
adding misc.forsale.txt
adding rec.motorcycles.txt
adding alt.atheism.txt
adding sci.crypt.txt
adding comp.os.ms-windows.misc.txt
adding talk.politics.misc.txt
adding comp.sys.mac.hardware.txt
.commit index None
done
0:00:05.035237
```

Once the index file has been created, you can see the message “done”

You can find the “Index.index” file in the same directory as your code. (If you can’t see the index, refresh the explorer)



2. Now that the index has been created, you can run a search query for the files. In the terminal, type: `python /home/Lucene/Code/SearchFiles.py`
Note that the query is case insensitive.

```

root@5e62a4ed2c7d:/# python /home/Lucene/Code/SearchFiles.py

Press enter with no input to quit.
If field is blank, there is no data for it.
Query: crypto
Searching for: crypto
3 total matching documents found.

Link to file /home/Lucene/Data/sci.crypt.txt

Link to file /home/Lucene/Data/alt.atheism.txt

Link to file /home/Lucene/Data/soc.religion.christian.txt

Press enter with no input to quit.
If field is blank, there is no data for it.
Query:

```

Since the above code returns only the link, we partitioned the 20 files into about 42,000 articles (partitioned on the Newsgroup line.) This will help us retrieve more data about the file, and be able to point to the specific article.

3. In the terminal, type: `python /home/Lucene/Code/FieldIndex.py`
`/home/Lucene/DataNews`

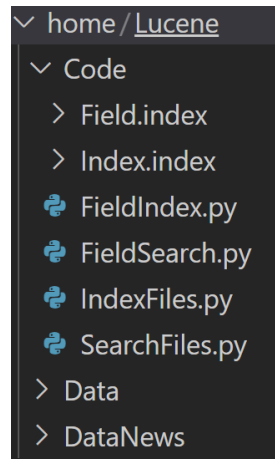
```

adding alt.atheism_4359.txt
adding comp.graphics_1869.txt
adding alt.atheism_4474.txt
adding comp.graphics_1745.txt
adding talk.religion.misc_1013.txt
adding talk.politics.guns_949.txt
adding comp.graphics_2048.txt
adding alt.atheism_1843.txt
adding alt.atheism_3834.txt
WARNING: No content in alt.atheism_3834.txt
.Commit index None
.Done
0:00:13.641115

```

Once the index file has been created, you can see the message “done”

You can find the “Field.index” file in the same directory as your code. (If you can’t see the index, refresh the explorer)



4. Now that the index has been created, you can run a search query for the files. In the terminal, type: `python /home/Lucene/Code/FieldSearch.py`

```
root@5e62a4ed2c7d:/# python /home/Lucene/Code/FieldSearch.py

Press enter with no input to quit.
If field is blank, there is no data for it.
Query: crypto
Searching for: crypto
25 total matching documents found.

Link to file: /home/Lucene/DataNews/sci.crypt_447.txt
Document_ID: 15598
From: strnlght@netcom.com (David Sternlight)
Subject: Re: An Open Letter to Mr. Clinton

Link to file: /home/Lucene/DataNews/sci.crypt_301.txt
Document_ID: 15449
From: <U23590@uicvm.uic.edu>
Subject: Cripple Chip

Link to file: /home/Lucene/DataNews/sci.crypt_1052.txt
Document_ID: 15203
From: lbrintle@news.weeg.uiowa.edu (Lee Brintle)
Subject: Re: Re-inventing Crypto Policy? An EFF Statement
```

Note: We have used the top 50 documents. This means we can obtain a maximum of 50 results.
To open the file, you can hover your mouse over the “Link to file:” and it will give you the option to open the file in the editor.

```
root@5e62a4ed2c7d:/# python /home/Lucene/Code/FieldSearch.py
```

Press enter with no input to quit.

If field is blank, there is no data for it.

Query: crypto

Searching for: crypto

25 total matches

[Open file in editor \(ctrl + click\)](#)

Link to file: /home/Lucene/DataNews/sci.crypt_447.txt

Document_ID: 15598

From: strnlght@netcom.com (David Sternlight)

Subject: Re: An Open Letter to Mr. Clinton