

Write-up \LaTeX File

Angatvir Sanghera, Tyler Crabtree, Tyler Coy

March 13, 2017

Abstract

Assignment 3 required us to create an inverted index on our databases which we are able to search through using Whoosh.

1 Attached Files Summary

We have attached all code files remember this was done in Python3. We have also given the SQLite3 database files which were used to create the index. We have also included segment files which are in indexdir.

1.1 Code File

There are two code files one of the files is indexer.py This first file takes in the databases and parses our databases and creates an inverted index over it. The second file which is searchEngine.py opens up the documents and does a search over it outputting the results found.

2 Team Name

Dragon Police

3 Database Schema

All our database schemas are completely different from one another. Our indexer.py gives the option of declaring our schema's name or going with the default. If default is chosen then the keys are grabbed from the database. We did it this way just in case we wanted to include other databases later. Each of default databases schema's is shown below.

3.1 Tyler Coy's Schema - dinosaur.db

Schema(Name=TEXT(stored=True), Description=TEXT(stored=True), Era=TEXT(stored=True), Url=TEXT(stored=True), Image=TEXT(stored=True))

3.2 Angatvir Sanghera's Schema - mmorpg.db

Schema(Title=TEXT(stored=True), Status=TEXT(stored=True), Graphics=TEXT(stored=True), Setting=TEXT(stored=True), SubscriptionModel=TEXT(stored=True), InitialReleaseDate=TEXT(stored=True), ServerCloseDate=TEXT(stored=True), Notes=TEXT(stored=True), Description=TEXT(stored=True))

3.3 Tyler Crabtree's Schema - superfamicom.db

Schema(Title=TEXT(stored=True), ReleaseDate=TEXT(stored=True),
Developer=TEXT(stored=True), Publisher=TEXT(stored=True))

3.4 Example of choice given

```
Creating new index: indexdir

Index: dinosaur.db already exists in directory: indexdir Overwrite?. YES or NO
<> YES

DATABASE: dinosaur.db contains 1 tables
Indexing table: dinosaur...
> You didn't provide a schema. Provide one? Yes or no (Use default schema): no
Proceeding to use default schema: Schema(Name=TEXT(stored=True), Description=TEXT(stored=True), Era=
TEXT(stored=True), Url=TEXT(stored=True), Image=TEXT(stored=True))
Indexing dinosaur.db
Total data indexed: 600

Index: mmorpg.db already exists in directory: indexdir Overwrite?. YES or NO
<> YES

DATABASE: mmorpg.db contains 1 tables
Indexing table: games...
> You didn't provide a schema. Provide one? Yes or no (Use default schema): no
Proceeding to use default schema: Schema(Title=TEXT(stored=True), Status=TEXT(stored=True), Graphics
=TEXT(stored=True), Setting=TEXT(stored=True), SubscriptionModel=TEXT(stored=True), InitialReleaseDa
te=TEXT(stored=True), ServerCloseDate=TEXT(stored=True), Notes=TEXT(stored=True), Description=TEXT(s
tored=True))
Indexing mmorpg.db
Total data indexed: 196

Index: superfamicom.db already exists in directory: indexdir Overwrite?. YES or NO
<> YES

DATABASE: superfamicom.db contains 1 tables
Indexing table: famicom...
> You didn't provide a schema. Provide one? Yes or no (Use default schema): no
Proceeding to use default schema: Schema(Title=TEXT(stored=True), ReleaseDate=TEXT(stored=True), Dev
eloper=TEXT(stored=True), Publisher=TEXT(stored=True))
Indexing superfamicom.db
Total data indexed: 1444
```

4 Size of Database

1444

5 Searches

We made our searcher very interactive. You can either search by categories or all categories.

5.1 Search 1

Just searched the word hello in the category description

SEARCH MODES:

1. Search by category
 2. Search all categories
- Press ENTER to use default [Search by category]

<> 1

[SEARCH MODE] Choose category

The following categories are available

- 1 to search in [Description]
- 2 to search in [Era]
- 3 to search in [Image]
- 4 to search in [Name]
- 5 to search in [Url]
- 6 to search in [Description - mmorpg.db]
- 7 to search in [Graphics]
- 8 to search in [InitialReleaseDate]
- 9 to search in [Notes]
- 10 to search in [ServerCloseDate]
- 11 to search in [Setting]
- 12 to search in [Status]
- 13 to search in [SubscriptionModel]
- 14 to search in [Title]
- 15 to search in [Developer]
- 16 to search in [Publisher]
- 17 to search in [ReleaseDate]
- 18 to search in [Title - superfamicom.db]

<> 1

Please input a QUERY string to search for in the database

e.g. Jurassic, Jurassic AND Ceratosaurus OR Stegosaurus etc.

You can also append a <==> to the end of the string to specify amount of results to display

e.g. Jurassic <==> 10 i.e. Provide top 10 result of searching 'Jurassic'

You can press /QUIT

<> hello

1 TOTAL hits

1 found in: mmorpg.db

Displaying 1 - 1 of 1 results.

1. <Hit {'Description': 'Hello Kitty Online (HKO) is a free-to-play[7] massively multiplayer online role playing game developed by Sanrio Digital and Typhoon Games.[8] The game is tied to the SanrioTown website, where players are required to register a free account.[9]', 'Graphics': '3D', 'InitialReleaseDate': '2009', 'Notes': 'Client download required.', 'ServerCloseDate': '', 'Setting': 'Cartoon', 'Status': '', 'SubscriptionModel': 'Free-to-play', 'Title': 'Hello Kitty Online'}>

PRESS ANY KEY TO CONTINUE or SAVE to save results or DISCARD to discard results

5.2 Search 2

Just searched Jurassic limiting results to 5

3 searchables found...

SEARCH MODES:

1. Search by category
2. Search all categories

Press ENTER to use default [Search by category]

<> 2

[SEARCH MODE] All categories

Please input a QUERY string to search for in the database

e.g. Jurassic, Jurassic AND Ceratosaurus OR Stegosaurus etc.

You can also append a <==> to the end of the string to specify amount of results to display

e.g. Jurassic <==> 10 i.e. Provide top 10 result of searching 'Jurassic'

You can press /QUIT

<> Jurassic <==> 5

153 TOTAL hits

152 found in: dinosaur.db

1 found in: superfamicom.db

Displaying 1 - 5 of 153 results.

1. <Hit {'Description': 'Juratyran (meaning "Jurassic tyrant") is a tyrannosauroid dinosaur genus from the late Jurassic period (early Tithonian age) of England.', 'Era': 'Late Jurassic', 'Image': 'https://upload.wikimedia.org/wikipedia/commons/thumb/b/bf/Juratyran_signed.jpg/50px-Juratyran_signed.jpg', 'Name': 'Juratyran', 'Url': 'https://en.wikipedia.org/wiki/Juratyran'}>

2. <Hit {'Description': 'Archaeodontosaurus ("ancient-toothed lizard") is a genus of sauropod dinosaur from the Middle Jurassic.', 'Era': 'Middle Jurassic', 'Image': 'https://upload.wikimedia.org/wikipedia/commons/thumb/a/a0/Archaeodontosaurus_dent_fond.jpg/50px-Archaeodontosaurus_dent_fond.jpg', 'Name': 'Archaeodontosaurus', 'Url': 'https://en.wikipedia.org/wiki/Archaeodontosaurus'}>

3. <Hit {'Description': 'Brachiosaurus is a genus of sauropod dinosaur from the Jurassic Morrison Formation of North America.', 'Era': 'Late Jurassic', 'Image': 'https://upload.wikimedia.org/wikipedia/commons/thumb/4/4b/FMNH_Brachiosaurus.JPG/50px-FMNH_Brachiosaurus.JPG', 'Name': 'Brachiosaurus', 'Url': 'https://en.wikipedia.org/wiki/Brachiosaurus'}>

4. <Hit {'Description': 'Dubreuillosaurus is a genus of carnivorous dinosaur from the middle Jurassic Period. It is a megalosaurid theropod.', 'Era': 'Middle Jurassic', 'Image': 'https://upload.wikimedia.org/wikipedia/commons/thumb/f/f0/Dubreuillosaurus_skull.jpg/50px-Dubreuillosaurus_skull.jpg', 'Name': 'Dubreuillosaurus', 'Url': 'https://en.wikipedia.org/wiki/Dubreuillosaurus'}>

5. <Hit {'Description': 'Dashanpusaurus (meaning "Dashanpu lizard" after the township it was discovered in) is an extinct genus of sauropod dinosaur that lived during the middle of the Jurassic period.', 'Era': 'Jurassic', 'Image': 'There is no image link for this object', 'Name': 'Dashanpusaurus', 'Url': 'https://en.wikipedia.org/wiki/Dashanpusaurus'}>

PRESS ANY KEY TO CONTINUE or SAVE to save results or DISCARD to discard results