

Seth Bergen
Eric Hsueh
Mini-Project #2 Project Report

General Overview & User Guide:

1. Place your source file in the same directory as all of the other files.
2. Open a terminal in said directory
3. Run 'python3 Phase_1.py'
4. Enter the name of your source file (include the file extension)
5. Ensure both break.pl and Phase_2.sh are runnable on your system using the chmod command.
6. Run './Phase_2.sh'
7. Run 'python3 Phase_3.py'
8. Enter your query in the format detailed in the project spec.

Algorithm For Evaluating Queries:

First split up the input into individual queries.

For each query:

- If the query is a single term with no delimiter (:,<,>), call fullSearch() looking for the term in any of text, name, or location.
- If the query has 'date' before the delimiter, call doDates() looking for the date in time and parsing the output based on the delimiter
- If the query has 'text' before the delimiter, call fullSearch() looking for the term in text only.
- If the query has 'name' before the delimiter, call fullSearch() looking for the term in name only.
- If the query has 'location' before the delimiter, call fullSearch() looking for the term in location only.
- If none of the above, or the query has multiple delimiters, it is invalid.

Take the intersect of each query's result with that of the last, or the result itself if it was the first one.

fullSearch() returns the IDs of the tweets that match the given criteria.

Use the IDs to fetch the full tweet records from tw.idx

Use xmlToReadable() to get nicely formatted versions of each record

Print these.

Testing Strategy

Ran the first example to make sure the file splitter works

Ran the first example to make sure that the database is correct

Used DB_Dump to verify that the output was what we wanted

Ran the 10 example test first

Ran queries against those 10 examples

Queries ran: against all the terms, and the locations

Then ran multiple lines with multiple queries inclusive of terms and locations

Ran the 1000 example test case
Ran queries against those 1000 examples
Ran the 1000 ones in the lab computers
Ran queries for the 1000 on the lab machines

Group Work Break-Down Strategy

Finished parts 1 and 2 before Sunday April 2nd. Then worked on Part 3 together during April 2nd.

Method of coordination was through text messages and constant followups that things were being done.

Part 1 was assigned to Seth

Part 2 was assigned to Eric

Part 3 we worked on it together to finish it off

Hours spent:

Seth: 12 hours finished Part 1, Part 2, Part 3

Eric: 13 hours finished Part 2, Part 3