CMPUT 291 Mini-Project 2

Design Document

By Michael Harbidge, Nicholas Wielgus and Parham Golestaneh

Part A: Overview and User's Guide

Our project consists of two parts, build_database.sh and main.py. Build_database.sh will create database indexes from four given files, pterms.txt, rterms.txt, reviews.txt, and scores.txt. It will return four .idx files used by the Python portion of the project. Upon running the Python file, it will open a command prompt like window where the user can type any query or type "q" to quit. The queries must be structured like shown in the specification of the assignment, where a field may be specified with a specific value and the program will return every result in the database that matches the query. When q is typed, the program will quit.

Part B: Software Design

Our project was coded mainly in Python, while also making use of the given perl script in the first phase. The second phase consists of one python file, Main.py, which contains an algorithm for evaluating queries given to it. The algorithm works by dividing the given query into separate tokens stored in a dictionary relative to what index they must access, for example pterm or rterm. The algorithm then calls a result returning function, which looks at the tokens for each index and adds any matching results to a set, which is then printed once every token is evaluated.

Part C: Testing Strategies

Our testing strategy consisted of using the 10 given queries, and ensuring that each query returns the correct results. If there was a problem with executing a given query, we went through the python file to find the bug and corrected it.

Part D: Group Work Breakdown

Work between group members was not really formally divided for this project, and each member contributed to the project whenever they were able to. In the end all group members put in roughly the same amount of time working on the project. Below is an overview of the major components of each file and who they were written by.

build_database.sh

Sorting and perl script calling: Parham

Db_load commands: Michael

Main.py

Tokenizing portion of query evaluation: Parham

Command prompt loop: Parham

Returning results based on given tokens: Michael, Nicholas, Parham

Interaction with database files: Nicholas, Michael