School of Science/ Computer Science and IT

COSC2636 Big Data Management

Assignment 2

Semester 1, 2019

1 Introduction

This is an individual assignment, to be submitted electronically using the Canvas facility. A submission link will be enabled on Canvas closer to the submission date. It is due 23:59 Friday of Week 8, and contributes 20% towards the aggregate of 100 marks.

The objective of this assignment is to reinforce what you have learned in the lectures and tute/ lab sessions. Specifically, Assignment 1 is centered around the IEEE TKDE paper "INSPIRE: A Framework for Incremental Spatial Prefix Query Relaxation" to implement some types of queries and build a mini google map.

Code Skeleton and guidance on backend implementation (you can implement everything from scratch at your own preference, the skeleton and API provided is just to help simplify your job), while you need to implement the frontend by yourself, using Google Maps API:

1.1 Plagiarism

All code or other material that is not original must be fully credited. That is, any material that is copied or derived from another source must be clearly identified as such and the original author must be identified. Sometimes students assist each other with an assignment, but end up working together too closely, so that the students' separate solutions have significant parts in common; unless the solutions were developed independently, they are regarded as plagiarised.

Plagiarism is a very serious offence. Any submissions determined to be a result of plagiarism will be deemed as an academic misconduct and harsh penalties apply. It is also an offence for students to allow their work to be plagiarised by another student. You should familiarize yourself with the university website for Academic Integrity Policy, Procedures and Guidelines. (https://www.rmit.edu.au/students/student-essentials/rights-and-responsibilities/academic-integrity) All work is to be done individually and plagiarism of any form will be dealt with according to the RMIT plagiarism policy.

1.2 What to Submit, When, and How

1.2.1 When

This assignment is due at 23:59 Friday of Week 8.

1.2.2 What

Please submit a zip file, naming convention: StudentNumber_A2.zip

The zip file should contain

- a folder, storing all your codes and a readme.txt describing how to run your code
- a report in pdf format, naming convention "report_studentnumber.pdf".
At the beginning of the report, write down: student number and name.
.
Regarding what should be included in the rest of the report, refer to the next paragraph.

A report should include the implementation procedure. Besides, you can include anything that you think helpful and useful. Note that the report is used to assist in our evaluation while there is no credit for report, but everyone needs to do coding. There is no page limit or format specification for the report, so long as it serves the purpose to tell us how you implement the ideas.

1.2.3 How

You are required to submit your solution electronically using the Canvas facility. A submission link will be enabled on Canvas closer to the submission date.

1.2.4 Penalties for late submissions

Late submissions of assignments will be penalised as follows. For 1 to 5 days late, a penalty of 10% (i.e. 10% out of total marks, not 10% out of your marks) per day. For assignments more than 5 days late, 100% penalty applies.

1.2.5 Special Consideration

If unexpected circumstances affect your ability to complete the assignment you can apply for special consideration. If you seek a short extension, you can directly contact the lecturer. For longer extensions, you must follow instructions provided at http://www1.rmit.edu.au/students/specialconsideration

1.3 Preparation Tasks

The code skeleton is provided in Canvas. Use the skeleton to implement the three functions missing in the file “QueryEngine.java”

2 Task: Implement three query functions (15 Marks)

You are required to implement functions of three queries -- prefix range query, substring range query and approximate prefix range query. The code skeleton is provided in Canvas. These functions to be implemented locate from line 497 to 653 in "QueryEngine.java."

Once you have implemented these functions, you can execute your program with commands over "RunQueryLatest.java".

An exemplar command is: index query.txt index index 2 3 9 10 5 5 10 0.01 0.0025 2 0

We have provided sample query file and the corresponding correct output.
You can check the correctness of your program with these files provided in Canvas.

3 Live demonstration (5 marks) and Grading criteria

Live Demonstration time: see course syllabus (to be arranged in practical sessions)
It is recommended to bring your own laptop to demonstrate the system.
You need to build a web service for the whole assignment and we could run your web service directly, and you should have your user interface to present the query results, e.g. using google map view.
Demonstration requirement:
You need to prepare some running example for your system beforehand.
We will also pick some queries to test your system.
The system should display the running time and the choice of
- Result size threshold (10 in the paper)
- Edit distance threshold (20% of the query length for SAP and SAS queries in the paper)
- length of q-gram (suggested: 3)
- length of positional q-gram (suggested: 2)

Should you want more details, refer to Sec 7.1 of the paper.

Please note:
1. You can adjust the setting for the above result size and edit distance thresholds, according to your own need.
2. It is not necessarily to be exactly same as the paper's choice, because you are working on a very small dataset with shorter average string length.
3. You just need to provide in report your choice of those thresholds and q for your implementation, and tell us your choices when you do the demonstration to us.

We mainly grade your program based on the correctness and efficiency. The 5 marks for live demonstration will be based on your presentation and your answer to our questions to demonstrate that you are confident and clear about each step done for each step of the assignment.