Department of Computer Science and Engineering
The Hong Kong University of Science and Technology

CSIT6000P Spatial and Multimedia Databases

2023 Spring

Version 1.0

Group Project and Report [*Total: 40 marks*]

Due date: Various deadlines
HKUST Canvas online submission only.

[Note: As discussed during the lectures, we will also allow some groups to do an implementation-based project, using PostGIS/QGIS and HK open spatial datasets. If any group is interested in this type of project, please form your group and discuss with the lecturer and TA to develop a project specification and marking scheme for your project.]

In this course, we discuss some fundamental concepts of spatial and multimedia databases as well as general high-dimensional data management and processing. There are still many challenges and open issues in this area in terms of both efficiency and effectiveness, such as high dimensional indexing and search, query processing in road networks, spatial-keyword search, spatiotemporal data analytics, retrieval of multimedia and social media data, dimensionality reduction and approximate search, and stream query processing. Recently we also see advances in applying machine learning methods to extract features (e.g., for natural language processing and computer vision-related work) and to index data (e.g., the learn-to-index approaches). Research results on these topics can be found in many recent papers published at leading conferences.

This project requires students to form groups to investigate one chosen advanced topic by studying a cluster of related high-quality research papers to develop their own deep understanding and possibly new ideas on the topic. Your group is required to prepare a research report and make a presentation about your study.

Forming Groups

Students are asked to form a group of 5 students (any variation of group size must be approved by the lecturer or the tutor in writing). Students should first attempt to form groups of 5 students by **5pm on the 4th of March.** Once you have formed a group, an email should be sent to the tutor listing the **name and student number of each student** in your group. Students looking for groups or recruiting group members should do so at the weekly lectures or via Canvas.

Students who are not part of any group by this deadline will be randomly assigned to a group by the tutor. The lecturer and tutor reserve the right to adjust a group's size to 4 or 6. No changes to grouping are allowed after this point of time.

Choosing Your Topic

The proposed topic of appropriate scope and relevance needs to be approved by the teaching staff.

Each group must choose at least **three** research papers published on one advanced topic of your choice that is related to what we discuss in this course. The following list provides a good starting point for your topic selection:

- High dimensional indexing and nearest neighbor search,
- Query processing in road networks,
- Spatial-keyword search,
- Systems and prototypes for spatial or trajectory data management,
- Spatiotemporal data analytics (i.e., spatial trajectory analytics),
- Retrieval of multimedia and social media data,
- Dimensionality reduction and approximate search (e.g., LSH),
- Stream query processing,
- Learn-to-index.

Do not choose more than 5 papers. Too-many papers selected will disadvantage your group, as we will focus on the depth of your understanding not the breadth. The research papers must have been published as full research papers in one of the following 5 conferences: SIGMOD, PVLDB, ICDE, SIGKDD, and ACM Multimedia. At least one paper must be published in or after 2020.

You should get the approval from the teaching staff via email by **5pm on the 25th of March**.

Preparing Your Research Proposal

The research proposal is a document up to 4 pages long that contain:

1. A clear and concise explanation (around a paragraph long) of the chosen topic and the scope of your research report.
2. A brief description of each of your chosen research papers and how they are relevant to the chosen topic. Here you are expected to discuss more than just your selected papers in order to illustrate why these papers are selected.
3. A justification of why the chosen topic is of suitable complexity and importance and how it is relevant to spatial and multimedia databases.
4. An initial allocation/delegation of work to group members.

This proposal should be prepared carefully, based on extensive discussions among the team members and possibly also with the Lecturer/Tutor.

Note: Proceedings of these conferences can be found online (using Google or DBLP), and the full text of the papers can often be downloaded freely. Should a password be needed, you can (1) find an alternative source that does not require a password; (2) access from HKUST Library through the HKUST site license; and (3) talk to your Lecturer immediately after failing other options.

Your research proposal will be due by **5pm on the 1st of April**.

Submitting Your Final Research Report

The research report is a document up to 8 pages long that includes/discusses:

1. A clear and concise explanation of the problem.
2. Application scenarios for the problem.
3. Related work prior to the chosen research papers.
4. New solutions proposed in the papers and their relationship with each other as well as other related works.
5. Methods used for evaluating the proposed solutions and experimental results.
6. Potential issues remaining for that problem with a brief analysis.
7. References (which will include but beyond the selected papers your group chooses to study).

The deadline for submitting your report is **the 4th of May at 5pm.**

Making Presentation

Each group will be given 10 minutes for presentation plus 3 minutes for Q&A. All presentations will be made during our regular teaching time (i.e., **from 7:30 pm – 10:30 pm on the 4th of May, Thursday**). Your group will be informed of the presentation schedule one week before.

Your group should video record your 10-minute presentation and submit the video by **5pm 1st of May**. A copy of your PPT should also be submitted together. In case we have to do online presentations due to the pandemic and we encounter any unexpected technical difficulties with your scheduled presentation, the lecturer will play the recorded video.

Submitting Deliverables

All documents should be submitted in PDF format. Documents should be properly referenced using the IEEE referencing style. Your group information (group ID, a list of names and IDs for all group members) must be included on the cover page of your reports.

Only a single group member needs to submit these assessment items. The following subsections describe each assessment item in further depth.

Assessment

The same score will be given to all members of a group. A peer assessment form will be made available after the final assessment due date. Based on the completed peer assessment forms, some non-performing group members may receive a reduced score.

Research Proposal [5 marks]: description clarity of your selected task [1 mark], proper justification of the task [2 marks], the purposes and expected outcome of conducting the task [1 mark], and a feasible and well-designed project plan [1 mark].

Research Report [20 marks]: the report will be marked as a whole package. The completeness of your work (in comparison to the scope established in the research proposal, unless modified with written permission from Lecturer/Tutor) can affect the overall marks of your report. This report is expected to have a good general introduction to the problem and overall cohesion towards your topic [4 marks]; a clear, easy-to-read, and self-contained technical section [8 marks] to describe the problem definition and variations as studied in your selected papers, an outline of their key ideas in your own language and their relationship in the suite of the papers you study, as well as how evaluations are done in those papers to establish the superiority of their proposed approaches; and a findings session [5 marks] to outline your findings, conclusions and future work on your chosen topic. Additional marks [3 marks] will be given at the marker's discretion based on factors including but not limited to insights, report structure, readability, clarity, and references.

Presentation [15 marks]: presentation marks will be given at the marker's discretion based on factors including but not limited to technical insights demonstrated, Q&A quality, PPT quality, presentation coherence, clarity, and timing.

Please design the project conservatively and carefully manage your group work and time to ensure your team can complete this assignment on time. This project requires a significant amount of self-study effort. Please ensure that your group makes good weekly progresses. Each group is welcome to discuss with the Lecturer/Tutor when doing this assignment.

Academic Misconduct

All submitted assessment items should be your original work. Coping text from any sources without explicit acknowledgment of the sources is considered plagiarism, which could lead to zero marks for your research report or the entire project. A plagiarism check will be conducted for all submissions; we strongly recommend that each group does such a check before submission. You do not want your group's results adversely affected by a group member who copied text from papers or online sources. Students are expected to be familiar with HKUST student misconduct guidelines. Further questions regarding what is acceptable should be directed to course staff via the course's provided communication channels.

Use of ChatGPT for producing these reports is not allowed!