COMP90050, Winter term, 2022
Project Description
Report submission due date: **July 29, 11:59pm** Melbourne Time

1 Introduction

This **project forms 40% of your final marks**. The project is about creating **a survey on a recent topic in Databases as a presentation and a report**. Writing a survey is a cornerstone activity that we would like you to learn as well as teamwork. In our sector, this is an activity you would need to do regularly to keep up-to-date with developments and in most cases present to your supervisors/companies that you work for in the form of a white paper/position paper and a company presentation.

This is a group project of **4-5** students. **The topic needs to be chosen from the list provided at the end of this document.** Here are the details of the project components -

1. **Forming a group:** This is an intensive semester of only 4 teaching weeks, so you need to form a group by the end of week 1. **You need to create a group with your group members in the People section of Canvas.** Forming a group in People section of Canvas help https://community.canvaslms.com/docs/DOC-10516-421264913 (Links to an external site. Please make use of your tutorial time and the Ed discussion in Canvas to find group members. Students from different tutorial slots can be in the same group.

2. **Choosing a topic:** Choose a topic as a team from the list of hot topics in databases given below to start the project. There is no advantage of choosing one topic to the other. Just a discussion among team members about their liking of an area is all that is needed to choose a topic. Then, **by the end of week 1 (July 3, 2022 11:59pm), you need to submit your group information and topic in Canvas->Assignments->A. Group topic submission.**

Submission format: Only one member of each group needs to make the submission. Enter text as the following format - Group ID (after joining in a group in the People section of Canvas you will see a group ID), Name and student ID of the group members, and the topic chosen by the group. The exact formatting of the text is not important as long as all these information are provided. If the group member information or the topic of your group changes, you may re-submit any time before your presentation. However, the first submission needs to be made by the due date **July 3, 2022 11:59pm.**

3. **Starting the work and expectations:** For your topic of choice, you should initially start reading Wikipedia, newspaper/magazine, and similar articles/webpages to get a high-level idea for what the topic is about. You can do this even on day 1 of the semester and for any or all of the proposed topics to get an overall idea. These articles you check are not adequate for a survey though but good for a start.

After this initial phase, you should use **scholar.google.com or similar scholarly search engines** for performing a more detailed background search and do further reading. You should distribute the work evenly among team members. These specialized search engines give you the papers you will cite for your survey that will really count. Generally, highly cited papers

are more important ones in an area. You should use multiple keywords/searches to cover a topic to find all the related papers. (It is important to note that if you login to our library with your student credentials, you will be able to access papers and some books that are returned by these engines for free in many cases.)

You are encouraged to find other survey papers that already exist in your topic as well. You do not need to start your survey from scratch basically. There would be many surveys out there potentially. Find one that is most recent. Better, find many and you will see authors look at similar but not the same set of algorithms/papers. They may also have different classifications and organization of things. These should give you an idea on what common/popular methods exist and what key comparison parameters you can have between solutions. They are also good examples on how to write surveys. You cannot use other surveys or other papers at large to copy things directly into your own surveys! These should not be the sole source of your work anyway. After you read and refer to these, you need to go to individual key papers mentioned in the surveys and your searches and start reviewing them and form your own opinions i.e., you need to make your own judgements and categorizations and also check more recent works that may not appear in earlier surveys (although many of the categories you create would likely be similar to older survey papers of the topic/area.)

As expected from any survey, **you are expected to not only list top papers in an area one after the other, but also categorize these developments/approaches and compare/critique them**. This is at the core of a survey. A list of papers with comments attached is called an "annotated bibliography" and is not a survey and is not the purpose of this project.

It is important to note that we do not expect you to learn every paper in detail and to be perfectly comprehensive about the topic to cover all the papers. But rather cover the key papers/directions and classifications/parameters. Digest the key directions and ideas basically! The number of citations a paper gets in scholar.google.com is an indicator about its leadership in the field, i.e., in addition to the fact that it is mentioned in other surveys.

The stages of your project can be summarized as: Background search/reading selected papers; then organization of your report with titles and subtitles, figure captions, etc, comes… while populating sections with key bullet points/issues to mention in these sections… Finally, finishing your report by writing the details of each point and drawing figures/tables. You should prepare your presentation in tandem with your report that pretty much presents the key points in the report. After this exercise students are expected to have a good idea in the topic and be able to answer questions during their presentation.

Following **report section-headings/structure needs to be followed (with recommended approximate page counts per section mentioned and that aspect could vary to a degree)**:

- Identification info for students/title/abstract/etc (1 page cover basically)
- Introduction to the Topic Area (2 pages)
- Related Work Details (papers covered explained in brief in some structured manner) (7 pages)
- Comparison of Key Approaches/Papers (benefits and disadvantages from various aspects) (2 pages)
- Conclusions/Discussions and Future Directions (2 pages)
- References (1 page)

When reading the papers, please note that a technical paper is not read like a novel, i.e., not read from cover to cover sequentially or slowly, but is read in a manner that you can quickly grasp the key ideas, benefits/disadvantages. (Although technical papers can be read to the very detail or even one can contact authors for implementations, we do not need that level of reading per method/paper from you for your project/survey.)

4. **Presentation: The presentation will be done as a team (with as many members presenting as possible) at the end of the semester based on a schedule announced by the lecturer closer to the dates.** Given the dual delivery mode, the presentations will be done as Zoom sessions. Tutorial slots of the last teaching week will be used to schedule the presentations. Structure of the presentation should follow the report as well, i.e., in terms of the key sections it involves. The presentations will be allocated **no more than 15 mins** per group including questions and setup time etc. Thus, we recommend no more than 12 slides for presentations. **The presentation file needs to be submitted before your presentation in Canvas->Assignments->B. Group project presentation as a ppt, pptx, or pdf file**.

5. **Report submission:** The deadline of this project is July 29, 2022 11:59 pm. Late submissions will get a penalty of 10% per day. **The report must be submitted as a PDF file on LMS->Assignments->C. Group project report.** Handwritten reports are not accepted in any form. The report should be submitted by only one person in your group, but all students' IDs should be visible in the cover page.

The report should be in **A4 size paper in 11-point Times New Roman font** for the main text with **1.5 line spacing** with **1-inch margins**. It should **be single column**. The report **should not exceed 15 pages (but also no less than 10 pages)**. Figures and Table(s) (e.g., for comparing methods) is highly recommended and crucial in some cases. Text that describes methods is important to understand the algorithms. **Approximately 3500 words are adequate to cover a survey topic, but may vary based on the informative tables and figures that you use.**

All explanations should be your own words and proper citations should be used when needed. Teams should work independently of other teams and plagiarism as usual will be checked by markers and our systems. Not sharing information about papers you found is also important as finding papers is a part of the project. Submissions are checked in our systems with other submissions, including other subjects, offerings from different semesters, websites, papers in search engines, etc through a professional plagiarism checking system.

List of topics

Here we give a short list of hot **topics that you can choose from**. No other topics are acceptable. You need to focus on Database aspects of these topics as at times they relate to other areas of Computer Science as well. Topics that you can choose from are:

- Graph Data Management
- Stream Data Management
- Spatial Data Management
- Multimedia Data Management

Graph Data Management has become a hot area of research and development, especially with the increasing popularity of social networks. Stream Data Management has become popular with increasing availability of data in the form of a stream over networks, such as stock or news data coming in as a stream on the Internet. Spatial Data Management is important for traffic data, GPS traces of people's movement data, etc. Multimedia Data Management is popular with the increasing number of images and videos being generated every day.

Top conferences that you can find papers from on these topics in Database community are SIGMOD, VLDB, ICDE, SIGKDD, SIGIR, SIGSPATIAL, ICDM among others. There are also top journals we can recommend, TKDE, VLDBJ, ACM TODS. All of the publications that appear in these should pop up in scholar.google.com. Other top tier publication venues are also acceptable. There are numerous conference and journal rankings you can check for venue rankings in computer science. Starting reading papers from these venues would likely to expedite your progress.

4 Marking

The presentation constitutes 15% of your final marks. The **students in the same group will receive the same mark** for the presentation. Each presentation will be marked in three aspects:

- Knowledge: understanding of topic and comprehensiveness of discussion
- Delivery: clarity and engagement
- Teamwork: time management, flow, and distribution of workload

Each aspect will be **marked from 0 to 5**. The mark of your presentation will be calculated as the sum of the three aspects. Detailed marking criteria follows:

	0	**…**	**5**
Knowledge	• Does not cover any representative work related to the topic • Content is focused on a wrong topic		• Covers the representative publications/products related to the topic • Compares different approaches (methods/algorithms/products/etc.) • Covers the importance/potential of the existing work
Delivery	• Difficult to understand, e.g., full of technical jargons • Content is totally unorganized		• Slides and oral presentation are easy to understand for general audience • Presentation is clear and captures the audience

Teamwork	• Unbalanced presentation time between team members • Content from different team members is unrelated with each other • Time length of the whole presentation exceeds the limit		• Workload is well balanced between team members • Content from different team members tells a full story with fluid presentation flow • Time length of the whole presentation is well controlled • Can handle questions from the audience

The final report is 25% of your final marks. The marking is similar to presentation marking above in many ways.

	0	**…**	**5**
Knowledge Coverage	• Report is focused on a wrong topic • Does not cover any representative work related to the topic	…	• Shows a comprehensive survey of the work related to the topic • Presents representative works in the body of the report and in the references
Related Work Structure	• Papers are covered in a totally unorganized way • Development of the area is not visible to the reader	…	• Papers are well organized • Reader can see how the area has developed and/or covered by papers in subareas
Writing	• Contains many grammatical errors • Writing is difficult to understand • Writing style is not academic, e.g., using an informal tone	…	• Grammatically sound • Readers with minimal knowledge of the topic can understand the content • Uses academic writing style • Shows figures/diagrams that help readers understand the content

Format	• Format is not consistent with what is prescribed • Layout is awkward • Style of references is not consistent and academic across the reference section	…	• Format is consistent • Uses correct layout • All the references use the same proper style
Critical Analysis & Comparisons	• Plainly lists all the approaches without analysis	…	• Shows motivation, practical use and/or potential of the approaches • Categorises the approaches • Compares different approaches • Analysis from multiple perspectives

… End of Project Description..!