

Project

CSE6242-OAN Data and Visual Analytics

Instructor: Prof. Polo Chau

Grading & Schedule

1. [Proposal](#) (7.5% of course grade)
2. [Proposal presentation](#) (5%)
3. [Progress report](#) (5%)
4. [Final poster presentation](#) (7.5%)
5. [Final report](#) (25%)

See Course Schedule for all deliverables' due dates.

Important: you will be submitting multiple files as part of your project deliverables. We will deduct 5% from a project deliverable for **every file** whose filename or file format that is different from what we have specified. It is time consuming to find "missing" files or to guess their names.

For example, suppose the final report requires README.txt and report.pdf; if your team submit README.doc and report.doc, 10% will be deducted from the final report's score.

Teaming and Choosing a Topic

The work will be carried out in teams of **4-6 persons**. You are welcome to decide on who to team with, since each team needs to decide on their topic. You may want to try Piazza's "Search for Teammates" pinned post.

During the semester, we will open up a shared spreadsheet that all teams will enter their team member information.

● Pick your own topic:

- You need to justify that the topic is interesting, relevant to the course, and is of suitable difficulty.
- Required components:
 1. at least one **large, real dataset**;
 2. some **non-trivial** analysis/algorithms/computation performed on the dataset (e.g., computing basic statistics, like average, min/max will not be enough); and
 3. an **interactive user interface** that interact with the algorithms (can be visual, voice-controlled, on tablet, desktop, etc.).

● Harder way:

- Joint projects with other courses might be negotiable. You must obtain the instructor's approval, and you need to clarify exactly what will be done for this course that is on top of and different from what you will do for the other course.

- Projects related to your dissertation/master-project are also possible, as long as there is no '*double-dipping*', i.e., you clearly specify what the project will do, in *addition* to what you were planning to do for your thesis anyway.

Once you have selected a topic, you should do some background reading so that you are capable of describing, in some detail, what you expect to accomplish. For example, if you decide that you want to implement some new proposal for a multidimensional file structure, you will have to carefully read the paper that proposes similar structures, pinpoint their weaknesses, and explain how your approach will address these weaknesses. Once you have read up on your topic, you will be ready to write your proposal.

Proposal

Your proposal should answer Heilmeier's questions (all 9 of them; see list below); if you think a question is not very relevant, briefly explain why. In other words, your proposal should describe what you plan to do (the problem to address), why you want to do it, how you will do it (what tools? e.g., SQLite, PostgreSQL, Hadoop, Kinect, iPad, etc.), how your approach is better than the state of the art, why it may succeed, and when it does, what differences will it make, how you will measure success, how long it's gonna take, etc.

9 Heilmeier questions ([source](#))

1. What are you trying to do? Articulate your objectives using absolutely no jargon.
2. How is it done today; what are the limits of current practice?
3. What's new in your approach? Why will it be successful?
4. Who cares?
5. If you're successful, what difference and impact will it make, and how do you measure them (e.g., via user studies, experiments, ground truth data, etc.)?
6. What are the risks and payoffs?
7. How much will it cost?
8. How long will it take?
9. What are the midterm and final "exams" to check for success? How will progress be measured.

Your proposal should be fewer than **1000** words, excluding titles, section names, reference list, etc., but including the literature survey. It should use **12pt font, typed in PDF format** (can be created using any software, e.g., LaTeX, Word), and with figures, tables, etc. whenever useful. It should be self-contained. For example, don't just say: "We plan to implement Smith's Foo-Tree data structure [Smith86], and we will study its performance." Instead, you should briefly review the key ideas in the references, and describe clearly the alternatives that you will be examining.

An appendix is for optional, non-essential information. We may not read or even grade it.

Some teams, especially those that want to turn their project into a research publication, use LaTeX for type formatting. If your team chooses to go this route, you may consider using tools like Git ([GT](#) [GitHub](#)) or Overleaf to work on the article collaboratively. For the LaTeX template, we suggest ACM's standard template (sigconf). You may need to increase the template's default font size to 12pt (e.g., by changing "`\defACM@fontsize{10pt}`" in the `acmart.cls`). You are also welcome to use other templates (e.g., IEEE, Springer).

How to write the survey without using too many words?

- See other articles' related work sections for inspiration, e.g., [Apolo paper](#)
- Multiple papers may share similar themes, use similar methods so they may be summarized and discussed together.
- Note that survey account for 60% of proposal's grade, so your survey should be substantial!

Grading scheme & Submission instructions

- 60% for the survey
- 30% for innovation
- 10% for plan of activities
- For every Heilmeier question that's **not** mentioned, deduct 5%.
- You may consider organizing your proposal based on the Heilmeier questions (e.g., each section addresses one question)
- Your survey should have **at least 3** papers or book chapters per group member (outside of any required reading for the class).
 - Short papers, like PNAS, Nature, Science papers, count as 0.5.
 - Copying the abstract of the papers is obviously prohibited, constituting plagiarism.
 - For each paper, describe
 - (a) the main idea,
 - (b) why (or why not) it will be useful for your project, and
 - (c) its potential shortcomings, that you will try to improve upon.
- You may use any citation style (e.g., APA, Chicago). Google Scholar supports a wide range of citation styles; it also provides BibTeX (needed if your team is using LaTeX).
- **Clear problem definition:** give a precise formal problem definition, in addition to a jargon-free version (for Heilmeier question #1).
- Provide a **plan** of activities and time estimates, per group member. **List what each group member has done, and will do.**
- Team's contact person submits a softcopy, named **teamXXproposal.pdf**, via Canvas (i.e., that person submits for the whole team), where XX is the team number (e.g., team01proposal.pdf for team 1)
- [-5% if not included] Distribution of team member effort. Can be as simple as "all team members have contributed similar amount of effort". If effort distribution is too uneven, I may assign higher scores to members who have contributed more.

Proposal Presentation

- Team's contact person submits
 1. A presentation slide deck via Canvas, called **teamXXslides.pdf**, where XX is the team number (e.g., team01slides.pdf for team 1). **PDF only**; no PPT or other formats.
 2. A **2-minute** video presentation (one presentation per team), called **teamXXproposal.mp4** (or .avi or .mov), where XX is the team number (e.g., 01 for team 1).
- Your video should show your slides (e.g., as pdf on your computer screen via screen capture, say using Quicktime, MonoSnap, etc.) with voice narration; it is up to you whether to show your face. You should be able to create this recording quickly with little effort – no need to do any special video or audio editing.

Grading

- [45%] You must answer the Heilmeier questions. 5% for each question. If a question doesn't apply, say so.
- [15%] Brief literature survey. Can be combined with Heilmeier question(s).
- [10%] Expected innovation. Can be combined with Heilmeier question(s).
- [10%] Plan of activities
- [20%] Presentation delivery
- [-5%] Illegible text, tiny figures, bad color contrast, etc.
- [-5%] Overrun
- Your presentation does NOT need to strictly follow your project proposal document. For example, you can talk about ideas and materials that your team has come up recently.
- **Points will NOT be deducted or awarded based on the number of presenters.** We saw great presentations delivered by teams having various numbers of presenters.

Tips

- Use few slides. Less is more! Fewer slides mean less likely to overrun. Being succinct is hard.
- Practice timing and delivery! If you have several speakers, make sure you practice how to transition from one person to the next (e.g., passing the mic, passing control of mouse and keyboard, etc). **PRACTICE! PRACTICE! PRACTICE!**

Progress Report

This should be fewer than **1600 words, 12pt font, typed**.

It mainly serves as a **checkpoint**, to detect and prevent dead-ends and other problems early on.

It should consist of the same sections as your final report (introduction, survey, etc), with a few sections "under construction", describing the work performed up to then, and the revised plans for the whole project.

Specifically, the introduction and survey sections should be in their final form. The section on the proposed method should be almost finished. The sections about experiments and conclusions will have whatever results you have obtained, as well as "place-holders" for the results you plan/hope to obtain.

The progress report may be written based on your proposal. For example, the survey in the progress report is not required to be identical to the survey in the proposal. That is, you may update the proposal's survey as needed. Of course, the number of papers should not drop below the requirement (3 papers/team member), and the quality of discussion should still be equal or better than that in the proposal.

An appendix is for optional, non-essential information. We may not read or even grade it.

Grading scheme & Submission instructions

- [70%] for proposed method (should be almost finished)
- [25%] for the design of upcoming experiments / evaluation
- [5%] for plan of activities (please show the old one and the revised one, along with the activities of each group member)
- Clear **list of innovations**: give a list of the best 2-4 ideas that your approach exhibits.
- Team's contact person submits a softcopy via Canvas (progress report only), named **teamXXprogress.pdf**, where XX is the team number (e.g., team01progress.pdf for team 1)
- [-5% if not included] Distribution of team member effort. Can be as simple as "all team members have contributed similar amount of effort". If effort distribution is too uneven, I may assign higher scores to members who have contributed more.

Final Poster Presentation

Overview

- Each team will create a single poster (for the whole team).
- Each team member submit, separately, an individual **3-minute** video presentation (one presentation per learner) via Canvas. Name each video recording **teamXXposter-YY.mp4** (or .avi or .mov), where XX is the team number (e.g., 01 for team 1), and YY is the student's last name (e.g., smith). Student A would submit A's video, and student B would submit B's video.
- Thus, every team member should know his/her project very well. Each team member should plan his/her presentation separately, and team members should not share presentation scripts.
- Each learner will grade several other video presentations given by learners from other teams.
- **Peer grading is NOT anonymous.** That is, a presenter knows who the graders are, and a grader knows who the presenters are.
- Your video should show your poster (e.g., as pdf on your computer screen via screen capture, say using Quicktime, MonoSnap, etc.) with voice narration; it is up to you whether to show your face. You should be able to create this recording quickly with little effort – no need to do any special video or audio editing.
- Demo: Optional but encouraged. The demo time counts towards the presentation time.

Why peer grading is not anonymous?

Polo wants students to learn and practice delivering constructive criticism, for any concerns and weaknesses identified.

People rarely like to hear about negative comments, even if they are accurate and helpful. Giving negative news is always hard, but that is part of life! This means we should carefully phrase our comments as constructive criticism. For example,

- instead of saying “too much text and not enough figures”, you could say “Fig 1 to 3 are important figures in this project; currently they are not easy to see (images are too small; text is not legible). Suggest reducing the amount of text, e.g., into succinct, bullet points to create space for the figures”.
- Similarly, avoid “I don’t think that the visualization is anything new or how it is helpful,” which is highly subjective. Instead, justify your comments; if the presenter did not clarify the novelty or significance of an approach (it is probably new, but just that the presenter did not point it out), you could say “it’s unclear from the presentation and poster whether the proposed visualization is an improvement over the state of the art (it seems to be a standard design); more clarification is needed.”

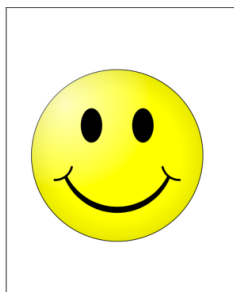
There are pros and cons for both anonymous and open review. It is still an open research problem. For example, one potential benefit for open review is reviewers could be more tactful and constructive, where anonymous reviewers could be more critical (sometimes not in a good way) and may do less work than they should.

Poster Design

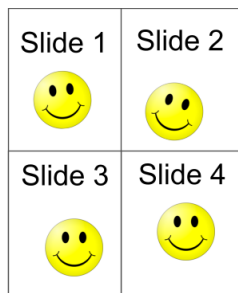
Design your team’s poster ***well before*** the submission deadline, to avoid last-minute rush.

The poster must be in portrait orientation, **30 inches wide and 40 inches tall** (if printed). We suggest using **18pt** font size and larger.

A deck of PowerPoint slides is **not acceptable** as a poster. See the illustration below for what is allowed and what is not.



OK



NOT OK

Grading

Your poster presentation should cover the following parts (point distribution shown on the left). Thus, the grading is about both your presentation delivery (e.g., what you say, where you direct the audience’s attention), and the poster content.

10%

Motivation/Introduction:

5% What is the problem (no jargon)?

	5% Why is it important and why should we care?
20%	Your approaches (algorithm and interactive visualization): 5% What are they? 5% How do they work? 5% Why do you think they can effectively solve your problem (i.e., what is the intuition behind your approaches)? 5% What is new in your approaches?
10%	Data: 5% How did you get it? (Download? Scrape?) 5% What are its characteristics (e.g., size on disk, # of records, temporal or not, etc.)
25%	Experiments and results: 5% How did you evaluate your approaches? 10% What are the results? 10% How do you methods compare to other methods?
10%	Presentation delivery: 5% Finished on time? 5% Spoke clearly and at a good pace?
25%	Poster Design: 5% Layout/organization (Clear headings? Easy to follow?) 5% Use of text (Succinct or verbose?) 5% Use of graphics (Are they relevant? Do they help you better understand the project's approaches and ideas?) 5% Legibility (Is the text and figures too small?) 5% Grammar and spelling

Possible software to create posters

1. Powerpoint/Word (save as pdf) -- GT's Office365 Powerpoint supports collaboration.
2. Apple Pages (FREE) supports real-time collaboration (via iCloud and desktop software)
3. Inkscape (free, cross platform)
4. Polo uses Affinity Designer (Mac and windows)

Example posters

- [Apolo graph exploration](#)
- [Insider trading pattern discovery](#)
- [Comment spam detection](#)

Final Report

It will be a detailed description of what you did, what results you obtained, and what you have learned and/or can conclude from your work.

Components:

1. **Writeup:** no more than **2800 words, 12pt font, typed**. Describe in depth the novelties of your approach and your discoveries/insights/experiments, etc.
2. **Software:** packaging, documentation, and portability. The goal is to provide enough material, so that other people can use it and continue your work, if you are to open-source it --- in other words, you should make it easy and attractive for others to use your work.

Grading scheme & Submission instructions

- Writeup

- [2%] Introduction - Motivation
- [3%] Problem definition
- [5%] Survey
- Proposed method
 1. [10%] Intuition - why should it be better than the state of the art?
 2. [35%] Description of your approaches: algorithms, user interfaces, etc.
- Experiments/ Evaluation
 1. [5%] Description of your testbed; list of questions your experiments are designed to answer
 2. [25%] Details of the experiments; observations (as **many** as you can!)
- [5%] Conclusions and discussion
- [-5% if not included] Distribution of team member effort. Can be as simple as "all team members have contributed similar amount of effort". If effort distribution is too uneven, I may assign higher scores to members who have contributed more.
- [10%] Team's contact person submits one zip file, called **teamXXfinal.zip**, via Canvas, where XX is the team number (e.g., team01final.zip for team 1). The teamXXfinal.zip will contain the following 3 components:
 - **README.txt** - a concise, short README.txt file, corresponding to the "user guide". This file should contain:
 1. DESCRIPTION - Describe the package in a few paragraphs.
 2. INSTALLATION - How to install and setup your code.
 3. EXECUTION - How to run a demo on your code.
 - **DOC** - a folder called DOC (short for "documentation") containing:
 1. **teamXXreport.pdf** - Your report writeup in PDF format; can be created using any software, e.g., latex, Word.
 2. **teamXXposter.pdf** - Your final poster.
 - **CODE** - All your code should be added here. Make sure that your package includes only the **absolutely necessary** set of files.

Should datasets be included as part of our submission?

If you are referring to (small) toy data for a demo (that your graders will run), you are welcome to include them. Think about the open-source software libraries that you have seen or have used, they would often include some sort of "quick start" guide to get a demo running on a toy dataset.

For large datasets, please do **not** include them; if the dataset is public and can be easily downloaded, include the link to the dataset.

If getting a dataset requires writing scripts/programs, include those scripts, and write down the steps that people will need to go through (e.g., register for an account to get API key).

If you have processed the dataset in some ways, include the code you used, and the steps people will need to go through.