Engaging Big Data | Spring 2016

Instructors: Catherine Murray, Fred Popowich, Peter Chow-White | Tuesdays 2:30–5:20 pm

Guest Speakers: Sasha Issenberg, Surajit Chaudhuri, Pat Hanrahan, Sheelagh Carpendale, Colin Hill, and Chad Skelton

This course is part of the SFU President’s Dream Colloquium on Engaging Big Data. Course application forms: www.sfu.ca/grad/events/dreamcolloquium/DreamColloquium-EngagingBigData.html

Description

This seminar explores the foundations and applications of “Big Data”. Given the increasing amount of digital information, many disciplines are bringing together data from different sources in an effort to generate new knowledge, weigh and mitigate risks, contribute more effectively to public policy and mobilize for positive social change. In a world which apparently digitizes everything but not everyone, there is a battle for access, and pressing need to develop the skills for analytic and interpretive reasoning, to mobilize big data in the public interest. The revealing of large private and secretive databases for marketing and national security has prompted public opposition to the risks of privacy. Some suggest big data dispenses with the need for theory. They argue we should let the big data speak for itself. However, others counter with a critique of this empiricist turn and pose critical questions for big data. They point out size may matter less than interpretive capacity, transparency, or the capacity of evidence-based decision making to adapt to the partisan inflection of every day relativist reasoning. The applicability of new analytic resources and processes is constrained by economic, legal and socio-political factors. The trick is to understand under what conditions and why this search for data-driven answers for societies most pressing issues may be effective in the next five to ten years.

The key questions that will be addressed during the seminar will be:
1. What is big data?
2. What can be done with it?
3. What should be done with it?

Within the context of these questions, this seminar intends to demystify the hitherto “privileged knowledge” of computing science, behind classification, detection and algorithms. This seminar will look at issues related the design and technology of big data, epistemological problems around datum evidence, analysis and meaning, ethics and regulation of its practices. The course mission is to seek to lever power from private to public spheres, through everyday applications. Cases will focus on how individual citizens, civil society organizations today can effectively access, interrogate and use large public databases to advance social, economic and political change. At the same time, we will be seeing examples of how Big Data and associated technologies pose risks for the “Little Guy”.

It is assumed that students will enter the course with an interest in a “wicked” research problem or persistent puzzle or paradox they want to explore by empirical means, related to their thesis or dissertation work, who wish to encounter others interested in the area and exchange knowledge, or brainstorm solutions. This course is a sandbox. It encourages students to take risks, access data they might otherwise not tackle, and produce something of use to someone or something somewhere. While many students taking the course may be in professional programs with corporate or government sponsors, the goal is to produce something of use to the not-for-profit, public knowledge dissemination sphere. Common intellectual property is assumed, with some exceptions granted under special circumstance. This course is envisioned as “pre-competitive” in stages of innovation, for those with interest in engineering or business.

Format

The class will meet weekly for three hours, with different instructors responsible for each week’s activity. Readings will be provided in-advance by the instructor or guest speaker for the week. During the weeks with a guest speaker, there will be a one hour presentation by the speaker, followed by discussions and a dinner. Students are expected to participate in the classes, invited-speaker seminars, and some of the dinners. During the term, students will work in groups on capstone projects that will be presented in-person and will be made available on-line at the end of the course. Note that some lectures might take place outside of the scheduled slot due to the schedules of the invited speakers. In weeks with local speakers, the seminar runs for two hours, followed by a workshop among students with presentations and discussion.

Evaluation

Students will be given an overall grade of satisfactory (pass) or unsatisfactory (fail). To obtain a satisfactory or passing grade, students will need to obtain a minimum of 70% in the course, based on the following activities with their associated weights.
• Weekly readings and associated written Exercises: 50%
• Participation: 10%
• Team Capstone Project: 40%