COMP 5117: Mining Software Repositories

Fall 2025
School of Computer Science
Carleton University

Class Schedule

Seminars are held every **Monday** from **11:35 AM to 2:25 PM** (online).
No classes on October 13 (Thanksgiving), October 20 (Fall break), November 10 (away at a conference).
Discord server will be used for course communication, news, and reminders. Please join the server here: https://discord.gg/SRPhCk4G98.
Schedule of seminars is posted on the course website.
Fall break: October 20–24, 2025 (no class).

Instructor

Dr. Olga Baysal
Email: olga.baysal@carleton.ca
Office Hours: via Zoom by appointment only
Website: http://olgabaysal.com/

Course Website

http://www.olgabaysal.com/teaching/fall25/comp5117_f25.html

Short Description

Introduction to the methods and techniques of mining software engineering data. Software repositories and their associated data. Data extraction and mining. Data analysis and interpretation (statistics, metrics, machine learning, AI, NLP). Empirical case studies.

Description

Software development projects generate impressive amounts of data. Mining software repositories research aims to extract information from the various artifacts produced during the evolution of a software system and inferring the relationships between them. This course will introduce the methods and tools of mining software repositories and artifacts used by software developers and researchers. Students will learn to extract and abstract data from software artifacts and repositories, such as source code, version control systems and revisions, issue-tracking systems, and mailing-lists and discussions. Students will also learn about various techniques of analyzing this data in order to identify meaningful relationships, patterns and trends, to recover behaviours and software development processes from evidence, or to empirically test hypotheses about software development.

Prerequisite

Students are expected to have some background in software development and software engineering. Prior knowledge of data mining, ML, AI, statistics, and natural language processing (NLP) would be an asset, but is not required.

Objectives

This graduate course explores the mining of data in software repositories in order to help researchers gain empirically based understanding of software development practices, and to support practitioners in managing, maintaining, and evolving complex software projects. The course will discuss leading research in the areas of mining software repositories. Papers discussed in this course will give students a glimpse of leading research which transforms software repositories from static record keeping repositories to active repositories that are used by researchers and practitioners to better understand and predict software development activities instead of depending on personal experiences and intuition. Students will be able to extract and analyze information from multiple software repositories in order to reason about existing software systems and development processes, as well to validate hypotheses about software development using data extracted from existing software systems.

Content Overview

The course will be adjusted according to students' interests and experience. This is an overview of the kinds of topics the course could cover:

- Mining software repositories
- AI for SE
- SE for AI
- Large language models (LLMs)
- Software development processes
- Software development tools and environments
- Software maintenance and evolution
- Collaborative development
- Human aspects in software engineering
- Quantitative and qualitative evaluation of software engineering research

Evaluation

- Weekly paper reviews: 10%
- Class participation and discussion: 20%
- Paper presentation: 10%
- Course project: 60% (10% project presentation + 50% project report)

Course Communication and Tech Support

- uOttawa OCICS students will not have access to Carleton Central. We will use Discord for course communication.
- University of Ottawa Students who need access to SCS IT resourcesm such as OpenStack and Nextcloud, must submit a request to SCS Tech Support at SCS.Tech.Support@cunet.carleton.ca. The request must be sent from their @cmail.carleton.ca email address and the email should say which resource is required and for which course (including section).
- Important dates and deadlines can be found here: https://carleton.ca/registrar/regulations/, including fall break and statutory holidays.

Mental Health and Wellness

https://wellness.carleton.ca/

Academic Accommodations and Regulations

For information about Carleton's academic year, including registration and withdrawal dates, see Carleton's Academic Calendar.

Academic Accommodations

Carleton is committed to providing academic accessibility for all individuals. Please review the academic accommodation available to students here: https://students.carleton.ca/course-outline/.

Academic Integrity

Students are expected to uphold the values of academic Integrity, which include fairness, honesty, trust, and responsibility. Examples of actions that compromise these values include but are not limited to plagiarism, accessing unauthorized sites for assignments or tests, unauthorized collaboration on assignments or exams, and using artificial intelligence tools such as ChatGPT when your assessment instructions say it is not permitted.

Misconduct in scholarly activity will not be tolerated and will result in consequences as outlined in Carleton University's Academic Integrity Policy. A list of standard sanctions in the Faculty of Science can be found here.

Additional details about this process can be found on the Faculty of Science Academic Integrity website. Students are expected to familiarize themselves with and abide by Carleton University's Academic Integrity Policy.

ChatGPT/Generative AI Usage

As our understanding of the uses of AI and its relationship to student work and academic integrity continue to evolve, students are required to discuss their use of AI in any circumstance not described here with the course instructor to ensure it supports the learning goals for the course.

AI use in this course (e.g., writing weekly paper reviews and project report): minimal use — basic assistance only. Use of AI is allowed only for automated grammar and punctuation checking tools (such as Grammarly).

Documenting AI use: It is not necessary to document the use of AI for the permitted purposes listed above. If you have questions about a specific use of AI that isn't listed above, please consult your instructor.

Student Rights & Responsibilities

Students are expected to act responsibly and engage respectfully with other students and members of the Carleton and the broader community. See the 7 Rights and Responsibilities Policy for details regarding the expectations of non-academic behaviour of students. Those who participate with another student in the commission of an infraction of this Policy will also be held liable for their actions.

Student Concerns

If you have any concerns regarding this course, your first point of contact is me. Please email me or visit during my student hours, and I will do my best to address your concerns. If I cannot resolve the issue, the next point of contact is the School of Computer Science at studentconcerns@scs.carleton.ca. If the concern remains unresolved, the final point of contact is the Office of the Dean of Science at ODScience@carleton.ca. Please follow this order of contact.

Note: You can also bring your concerns to Ombuds services.