# COMP 5117: Mining Software Repositories

# Fall 2024

# School of Computer Science Carleton University

#### Class Schedule

Seminars are held every Monday from 10:05 AM to 12:55 PM (online).

No classes on October 7 (I am away to a conference), October 14 (Thanksgiving), October 21 (Fall break).

Discord server will be used for course communication, news, and reminders. Please join the server

here: https://discord.gg/gpDkeVqJ7s.

Schedule of seminars is posted on the course website.

Fall break: October 21–25, 2024 (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/fall24/comp5117\_f24.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, machine learning, 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
- Large language models (LLMs)
- Software development processes
- Software development tools and environments
- Software analytics
- Software maintenance and evolution
- Sentiment analysis
- Collaborative development
- 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)

### University of Ottawa Student Access to Brightspace

- Please see information here:

https://gradstudents.carleton.ca/faculty-of-graduate-and-postdoctoral-affairs-access-to-b

- 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 resources 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).

## University Policies

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**

Student Academic Integrity Policy. Every student should be familiar with the Carleton University Student Academic Integrity policy. A student found in violation of academic integrity standards may be sanctioned with penalties which range from a reprimand to receiving a grade of F in the course, or even being suspended or expelled from the University. Examples of punishable offences include plagiarism and unauthorized collaboration. Any such reported offences will be reviewed by the office of the Dean of Science.

The use of any AI system will be considered academic misconduct. This includes, but is not limited to, chatbots or code generators (e.g., ChatGPT) for projects (reports, models, etc.). An exception to the above rule is made for automated grammar and punctuation checking tools (such as Grammarly).

More information on this policy may be found on the ODS Academic Integrity page: https://science.carleton.ca/students/academic-integrity/.

#### Plagiarism

As defined by Senate, "plagiarism is presenting, whether intentional or not, the ideas, expression of ideas or work of others as one's own". Such reported offences will be reviewed by the office of the Dean of Science. More information and standard sanction guidelines can be found here: https://science.carleton.ca/students/academic-integrity/. Please note that content generated by an unauthorized A.I.-based tool is considered plagiarized material.

#### **Unauthorized Collaboration**

Senate policy states that "to ensure fairness and equity in assessment of term work, students shall not co-operate or collaborate in the completion of an academic assignment, in whole or in part, when the instructor has indicated that the assignment is to be completed on an individual basis".