COMP0104: Software Development Practice
Coursework 2 (Group Coursework)

Notes

- Coursework 2: Collaborative, Summative Assessment.
- This coursework is a compulsory part of the module assessment.
  It contributes 15% of the overall marks for the module.
- The coursework is to be submitted online through Moodle
  by **Wednesday, 5 January, 2022 before 4pm**.
- There should be one submission per group. The submission should contain:
  - A PDF file with the report.
  - A ZIP file with all additional information and sources.
- The coursework will be marked anonymously as much as possible.
- You need to submit an ethics review request online through Moodle
  by **Tuesday, 7 December, 2021 before 4pm**.

Academic Support If you require support, please make use of this module's Moodle discussion forum in the first instance, which will be monitored by myself and the module's Teaching Assistants. You may also request an appointment to see me. Please note that my availability will be limited in this period.

Extenuating Circumstances If you become ill or experience any other sudden unexpected personal circumstances that affect your ability to undertake these assessments, please ensure that you submit a claim for Extenuating Circumstances to your home department as soon as possible, and within one calendar week of the circumstance taking place. I am unable to grant extensions unless approved through the formal process.

Academic Misconduct Please review UCL's guide on plagiarism and ensure that you fully understand what is and is not acceptable practice. If in doubt, please ask via the Moodle discussion forum. All work submitted for this module will be reviewed for evidence of plagiarism as a matter of course, and any evidence of academic misconduct will be fully investigated.

Academic misconduct also covers work needing ethical approval where this is not sought.

Preface

The coursework complements what is being taught in the lectures. While the lectures present you what industrial and academic research has established, the coursework will require you to research a data set and establish findings about software engineering tools and processes. It will expose you to research methods of the software engineering area and will provide you the experience of applying some of the tools you learned in class to real-world software systems. Finally, you will experience writing in an academic style, which you will need for your summer or final year project.

Mine Repositories from the Apache Project

The Working Conference on Mining Software Repositories is an established conference focusing on how the data in software repositories can be analysed and exploited. Each year, the conference posts a Mining Challenge where a large data set is provided and the community is asked to apply their tools and techniques on the data set and report on their findings.

In Coursework 1, you suggested a topic or question that can be analysed with the data underlying this year's Mining Challenge. In Coursework 2, you will form a group of 4 or 5 students and actually analyse some software repositories from the Apache project.

You will **not** use the data from the Mining Challenge as it contains data that might raise ethics issues.

This is a group coursework and you have to submit all programs or scripts in the end together with a report.

Building a Group

You should distribute the work between all group members. Some will focus on the practical side, i.e. understanding the data, some will focus on the theoretical side, i.e. analysing related work, while others will focus on the empirical side, i.e. doing the statistical analysis.

Find a Good Research Question

The first step is to get together and discuss the individual proposals that you created for Coursework 1. Based on the discussion, you may pick one of the proposals and actually do the proposed work. However, you can also decide to do a different approach.

You are allowed to refine/change the topic, it does not have to be one from Coursework 1.

If you cannot find a good research question, you can use the one below:

In Test-Driven-Development, tests are written before the tested code. If a project adopted TDD, the git repository should reflect this. If a new class (file) is created, then the same or an earlier commit should also create a new test class (file).

If you adopt the above research question, then you will need to answer questions like the following:

1. How often is a test class (file) created (a) before, (b) after, or (c) in the same commit as a tested class (file)?
2. How does the size of a commit impact the results?
3. How can you link a test class (file) to a tested class (file)?

You are free to go beyond the above questions. Groups using the above research question will not be disadvantaged in marking compared to groups using a different research question.

Analysing the Repositories

You will not use a prepared dataset, but you have to analyse the repositories directly. We suggest that you use PyDriller, a Python library that can be found at `https://github.com/ishepard/pydriller`.

Additional Coursework Requirements

Anonymous Marking and Submissions

The coursework will be assessed anonymously as much as possible. Therefore it is important that you do not reveal any personal information in any report and you only use your group id (i.e. Group A, Group B, Group C, etc.) In case your project needs to apply for a UCL ethical approval exemption, we will need to give student names and therefore it is not possible for full anonymity and we cannot guarantee anonymity.

Ethical Considerations

You will need to consider the ethical implications of your analysis. Therefore, you need to read the article on ethics in mining software repositories and watch the corresponding video. In addition, you need to take the departmental ethics training, at least the films covering UCL policy/process.

- Gold, N. E., & Krinke, J. (2022). *Ethics in the mining of software repositories*. Empirical Software Engineering, 27:17. `https://doi.org/10.1007/s10664-021-10057-7`
- Presentation: `https://mediacentral.ucl.ac.uk/Play/44038`
- Ethics Training: `https://moodle.ucl.ac.uk/course/view.php?id=16849` (enrolment code: csethics1920)

The considerations should be part of your report in the form of an appendix after the list of references.

To comply with UCL regulations on ethics, you are not allowed to use the dataset as provided by the conference. Instead, you will need to analyse the git repositories of the Apache project directly without using data from any other source. Please consult the lecturer of the module, Jens Krinke, or the TA, Chizzy Meka, if you have any question or when you are in doubt whether your analysis is in line with the advise given above.

Ethical Approval

Before you start analysing the data, you will need to submit an ethics review request which is reviewed by the module leader and another expert in ethics. You need to obtain approval to proceed.

The ethics review request should be a single page document with the following information:

1. The group id (e.g. Group A, Group B etc.)
2. A short description of the proposed analysis.

3. A clear statement of the ethical implications of the analysis.
4. A description of the elements of the data set that you will create and analyse.

A template is available on Moodle.

The ethics review request should be submitted as soon as possible via Moodle, no later than 7 December 2021 at 4pm, and you will be notified if it is approved or not as quickly as possible.

The ethics review request will not be marked, but it may inform the ethical considerations appendix in your report.

Formatting

You have to follow the ACM formatting instructions, which are available at `https://www.acm.org/publications/proceedings-template`. LaTEX users must use:

```
\documentclass[sigconf,review,anonymous,nonacm]{acmart}
```

Your report papers must not exceed 4 pages plus 1 additional page only with references. An additional sixth page is for the ethical consideration.

Additional Notes

- Your report papers must not exceed 4 pages plus 1 additional page only with references and must conform to the MSR 2021 format and submission guidelines.
- Different to the actual MSR requirements, an additional sixth page is for the ethical consideration.