Resit exam 09-04-2024

2IMP40 Empirical Methods in Software Engineering 23/24 · 4 exercises · 100.0 points

1 Part 1

20.0 points · 5 questions

The first part of the exam consists of **five** multiple-choice questions. Each question has four answer options, and exactly one is correct. Each correctly answered question gives four points.

Text

a Fang et al. (2024) consider as innovative the software projects that recombine existing software libraries in novel ways. They analyse GitHub projects with over 50% of their source code files written in Python and query the complete commit activities for all projects before the end of 2021. The results show that higher levels of innovativeness are statistically associated with higher GitHub star counts, i.e., novelty begets popularity. At the same time, we find that controlling for project size, the more innovative projects tend to involve smaller teams of contributors, aswell as be at higher risk of becoming abandoned in the long term.

This study can be classified as

4.0 points · Multiple choice · 4 alternatives

○ A field experiment	0.0
○ A field study	0.0
○ A judgement study	0.0
◉ A sample study	4.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

b Please consider the below abstract and indicate what type of design science contribution this paper is.

The behavior of software that uses the Java Reflection API is fundamentally hard to predict by analyzing code. Only recent static analysis approaches can resolve reflection under unsound yet pragmatic assumptions. We survey what approaches exist and what their limitations are. We then analyze how realworld Java code uses the Reflection API, and how many Java projects contain code challenging state-of-the-art static analysis. Using a systematic literature review we collected and categorized all known methods of statically approximating reflective Java code. Next to this we constructed a representative corpus of Java systems and collected descriptive statistics of the usage of the Reflection API. We then applied an analysis on the abstract syntax trees of all source code to count code idioms which go beyond the limitation boundaries of static analysis approaches. The resulting data answers the research questions. The corpus, the tool and the results are openly available. We conclude that the need for unsound assumptions to resolve reflection is widely supported. In our corpus, reflection can not be ignored for 78% of the projects. Common challenges for analysis tools such as non-exceptional exceptions, programmatic f iltering meta objects, semantics of collections, and dynamic proxies, widely occur in the corpus. For Java software engineers prioritizing on robustness, we list tactics to obtain more easy to analyze reflection code, and for static analysis tool builders we provide a list of opportunities to have significant impact on real Java code.

4.0 points · Multiple choice · 4 alternatives

○ Problem Solution	0.0
◉ **Descriptive**	4.0
○ Solution Validation	0.0
○ Solution Design	0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

c Regression discontinuity design...

4.0 points · Multiple choice · 4 alternatives

◉	**aims to determine the causal effects of interventions.**	4.0
○	refers to the point in data collection when no additional issues or insights emerge from data and all relevant conceptual categories have been identified, explored, and exhausted.	0.0
○	is typically associated with interviews.	0.0
○	might lead to resentful demoralization.	0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

d In the Strauss-Corbinian Grounded Theory building one is expected to perform several analysis steps. These steps are

4.0 points · Multiple choice · 4 alternatives

○	basic memoing, theoretical sampling and saturation	0.0
◉	**open coding, axial coding and selective coding**	4.0
○	dependability, credibility, transferability and confirmability	0.0
○	triangulation, sequential exploratory and sequential explanatory	0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

e Tools designed to enrich data from software repositories

4.0 points · Multiple choice · 4 alternatives

- (●) **might underperform when applied outside of the context they have been designed for.** 4.0
- ( ) require retraining for the context of a specific study. 0.0

  Feedback
  Not all tools can be retrained.

- ( ) require substantial computational power. 0.0
- ( ) are necessary to address the mono-method bias. 0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

2 Part 2

40.0 points · 5 questions

The text included below summarizes an introduction to a recent empirical paper. This summary concludes with a research question. For this part of the exam, we ask you to sketch an empirical study that answers the listed research question. For scoping you can assume that this study will be conducted as part of a master thesis graduation project (30 ECTS). This part is divided into several sub-questions. In particular, we ask you to sketch several alternative study designs for the same research question and discuss the differences between these designs.

Context and Research Question: *In this study, we focus on the emotions experienced by software developers at the workplace. Consistently with previous research on developers' emotions during programming tasks, we operationalize emotions along continuous dimensions. Following Russel, we describe the emotion stimulus in terms of its (un)pleasantness, ranging from low to high valence, and level of activation, ranging from low to high arousal. Furthermore, we include consideration of dominance, that is a person's perception of being in control of a situation. A priori, one might have thought that developers, being human beings, should experience the entire range of emotions at the workplace. However, different professionals have been shown to experience and express different ranges of emotions while at work: e.g., Foster and Sayers reported about physiotherapists not experiencing calmness and serenity, which in our terms would correspond to high valence and low arousal. As such, we formulate our first research question as follows:*

RQ: What is the range of developers' emotions at the workplace?

Based on the above context you should describe three different data collection procedures: a study using **Mining Software Repositories**, **Interviews**, and a **Survey**. For each of the three, you should discuss the data you plan to collect, where and/or how you plan to collect it (sampling) and sketch the instruments you plan to use. While describing the three different studies you should try to be as specific as possible.

Text

a Describe the data-collection of a study that answers the research-question using **Mining Software Repositories.**

8.0 points · Open · 3/5 Page

+4 points
Sampling -- Full Points: The answer clearly describes what sampling strategy will be used, it is tailored towards the setting of the research question, and motivates why the sampling is appropriate given the setting.

+2 points
Sampling -- Half points: Two of the three criteria are met.

0 points
Sampling -- No Points: One or less than the criteria are met.

+4 points
Repository Mining -- Full Points: The answer sketches what data and how will be mined to answer the RQs. The answer aligns with the described sampling technique, sketches what information will be extracted, operationalizes constructs, and conforms to best repository mining practices.

+2 points
Repository Mining -- Half Points: Two of the four criteria are met.

0 points
Repository Mining -- No points: One or less than the criteria are met.

b Describe the data-collection of a study that answers the research-question using **Interviews.**

8.0 points · Open · 3/5 Page

+4 points
Sampling -- Full Points: The answer clearly describes what sampling strategy will be used, it is tailored towards the setting of the research question, and motivates why the sampling is appropriate given the setting.

+2 points
Sampling -- Half points: Two of the three criteria are met.

0 points
Sampling -- No Points: One or less than the criteria are met.

+4 points
Interviews -- Full Points: The answer sketches what interviews will be conducted to answer the RQs. The answer aligns with the described sampling technique, sketches what questions will be asked, operationalizes constructs, and conforms to best interview practices.

+2 points
Interviews -- Half Points: Two of the four criteria are met.

0 points
Interviews -- No points: One or less than the criteria are met.

c Describe the data-collection of a study that answers the research-question using a **Survey.**

8.0 points · Open · 3/5 Page

+4 points
Sampling -- Full Points: The answer clearly describes what sampling strategy will be used, it is tailored towards the setting of the research question, and motivates why the sampling is appropriate given the setting.

+2 points
Sampling -- Half points: Two of the three criteria are met.

0 points
Sampling -- No Points: One or less than the criteria are met.

+4 points
Survey -- Full Points: The answer sketches what survey will be conducted to answer the RQs. The answer aligns with the described sampling technique, sketches what questions will be asked, operationalizes constructs, and conforms to best practices on conducting surveys.

+2 points
Survey -- Half Points: Two of the four criteria are met.

0 points
Survey -- No points: One or less than the criteria are met.

d You have described three alternative data collection procedures (2a), (2b), and (2c). List the unique advantages and challenges associated with each data collection procedure.

8.0 points · Open · 9/20 Page

+1.34 points
Advantages of the **repository mining** approach are correctly identified.

+1.33 points
Challenges of the **repository mining** approach are correctly identified.

+1.33 points
Advantages of the **interviews** are correctly identified.

+1.34 points
Challenges of the **interviews** are correctly identified.

+1.33 points
Advantages of the **surveys** are correctly identified.

+1.33 points
Challenges of the **surveys** are correctly identified.

e Pick one data collection procedure, and argue why, based on the advantages and challenges of each procedure, you believe it is the most appropriate research procedure for the listed research question.

8.0 points · Open · 9/20 Page

+4 points
The argument for the most appropriate procedure refers to advantages and challenges listed in 2d).

+4 points
The argument is valid (i.e., does not contain logical flaws).

3 Part 3

40.0 points · 6 questions

For the third part we ask you to read the accompanying paper. Please read the paper and come-up with two viable and distinct threats to validity. For each threat to validity please use the answer fields below to describe and classify it. The threats you describe should describe **two different types of threats to validity** (So if the first threat to validity you describe belongs to the category external, the second threat you describe cannot belong to the category external).

You should describe your first threat to validity in 3.a, describe the conclusion it invalidates in 3.b and classify it in 3.c, and you should describe your second threat to validity in 3.d, describe the conclusion it invalidates in 3.e and classify it in 3.f.

Text

a Describe your first threat to validity.

9.0 points · Open · 7/20 Page

+9 points

The answer describes a threat to validity that could indeed threaten the validity of the study

+4 points

The answer describes a potentially viable threat to validity. However, the description is too vague, or is not correctly argued.

b Describe the findings of the paper that would be invalidated by the threat to validity, and describe how these findings would be invalidated.

9.0 points · Open · 9/20 Page

+9 points

The answer points to a specific conclusion of the study that might be impacted by the threat to validity.

+4 points

The conclusion might not be fully impacted by the threat, or the description of the conclusion that is impacted is too vague.

c Classify the above threat using the model of Wohlin et al.

2.0 points · Multiple choice · 4 alternatives

Conclusion validity	2.0
Construct validity	2.0
External validity	2.0
Internal validity	2.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

d Describe your second threat to validity.

9.0 points · Open · 7/20 Page

+9 points
The answer describes a threat to validity that could indeed threaten the validity of the study

+4 points
The answer describes a potentially viable threat to validity. However, the description is too vague, or is not correctly argued.

e Describe the findings of the paper that would be invalidated by the threat to validity, and describe how these findings would be invalidated.

9.0 points · Open · 9/20 Page

+9 points
The answer points to a specific conclusion of the study that might be impacted by the threat to validity.

+4 points
The conclusion might not be fully impacted by the threat, or the description of the conclusion that is impacted is too vague.

f Classify your second threat using the model of Wohlin et al.

2.0 points · Multiple choice · 4 alternatives

Conclusion validity	2.0
Construct validity	2.0
External validity	2.0
Internal validity	2.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

4 Extra Space

0.0 points · 1 question

Extra space. Please clearly indicate to what question your answer belongs.

Open · 2 4/5 Page