Mining Software Repositories

Organization

Prof. Dr. Steffen Herbold

steffen.herbold@uni-passau.de

Lukas Schulte

lukas.schulte@uni-passau.de

Administration

- Lecture
  - Tuesday, 10:15-11:45 o'clock
  - Room: Building N 12 – Room SR 008

- Exercises
  - Thursday, 14:15-15:45 o'clock and 16:15-17:45 o'clock
  - Room: Building HK 12 – Room SR 001

- Exam date 1:
  - Friday, July 28th, 14:00-16:00 o'clock
  - Room: Building WiWi – Room HS 5

- Exam date 2:
  - Thursday, October 12th, 14:00-16:00 o'clock
  - Room: Building AM – Room HS 10

Exercise

- Application of Concepts from the Lecture

- Weekly exercise
  - Starts on April 27th
  - You can freely decide to which slot you go

- Combination of theoretical exercises and practical application development
  - Can be solved alone or in groups

- Exact format will depend on number of (active) participants during the exercise
  - If there are only few participants:
    - Students/group will get individual feedback for their practical solutions ("Testat")
    - Theory tasks will afterwards be discussed with the whole group
  - If there are too many participants for the above:
    - Everything will be discussed as a group based on the student solutions

Exercises: What they are and are not

The exercises are …

- optional in the sense that you do not automatically fail if you do not do the exercise

- for giving feedback for your solutions and give you hints on how to solve stuff

- an opportunity to better learn the concepts and work on your own solutions

- for learning about the contents of the lecture

The exercises are not …

- optional in the sense that you can expect a good grade (or even passing) if you do not solve the exercises

- a basic programming tutorial or the place where you start working on the solutions

- a second lecture in which we present or give solutions

- for discussing administrative issues

Communication and materials

- Everything runs via StudIP
    - Lecture slides
    - Exercise sheets
    - Other announcements from the lecturers
    - Literature hints

- Make sure you are enrolled in both the lecture and the exercise

Learning goals

- Know the role of MSR for software engineering tools and evidence-based software engineering

- Understand how to collect data from repositories and how this data can be enriched

- Understand how to design MSR studies and tools

- Be able to apply all of the above

Prerequisites

- Some machine-learning knowledge required
  - Foundations in data science
    - Data cleaning, feature engineering, statistics, machine learning

  - Software engineering incl. processes and quality assurance
    - E.g., version control, issue tracking, quality assurance approaches

  - Good programming knowledge
    - Most will be Python
    - Should already know common data science libraries like pandas, matplotlib and scikit-learn
    - Should be able to work with new libraries and tools just based on their documentation

  - Knowledge about working with and setting up databases

Contents of the lecture

- MSR data collection
  - Data sources and collecting raw data
  - Enriching data with heuristics and manual validation
  - Software metrics
  - Developer social networks

- Applications of MSR
  - Understanding defects
  - Studying social aspects of software engineering
  - Visualizing code and its history
  - …

Questions?