Mining Github Repositories

Joakim Leed, jolee18@student.sdu.dk

OnlyFans Hansen, ohans21@student.sdu.dk

SpankBanko, emspa15@ßtudent.sdu.dk   Jectho, jetof20@student.sdu.dk

March 7, 2025

Introduction

- What field are we in

- What problems is the field facing

- Any other important points for contextualizing this project

- What are our contributions

- The reasons why a rich data set would benefit the field

- Research questions

- Our plan to answer these research questions

-

Design

- Challenges to solve the issues
- How we address these challenges
- Diagrams of our data stages
- Design of our resulting rich data matrix

Implementation

- Overview of the implementation steps
- Highlights constraints of our implementations (e.g we shouldn't scrape too often from github)
- Share experiences of discovered pitfalls in the implementation
- Briefly document that our implementation works and where the code/dockerfile can be found and how to execute it

Related works

- Brief summary of related works
- Draw similarities between the related works and our approahc
- Postulate how our data could improve the related works

Conclussion

- summarize:
    - the field and its issues
    - How we contribute with our data set
    - Shortcomings of our data set
    - Where to go next