Interactive Visualizations for Software Duplication

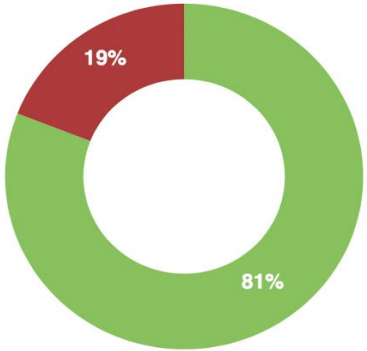
25.August 2020

Master Thesis of Jonas Richner

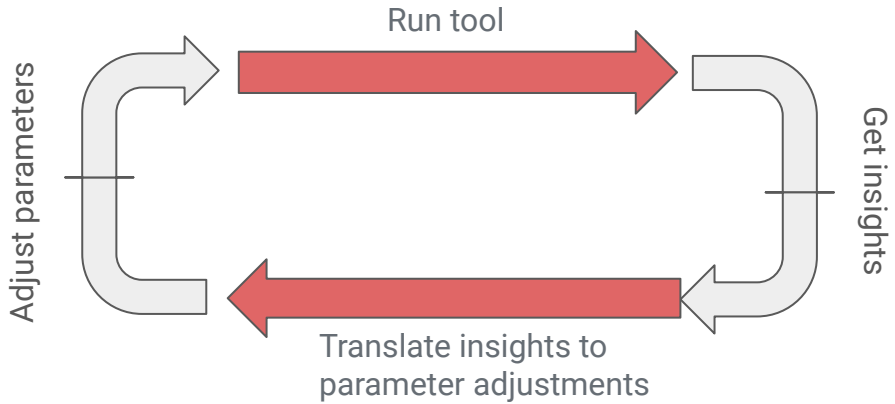
Duplicated code increases maintenance costs

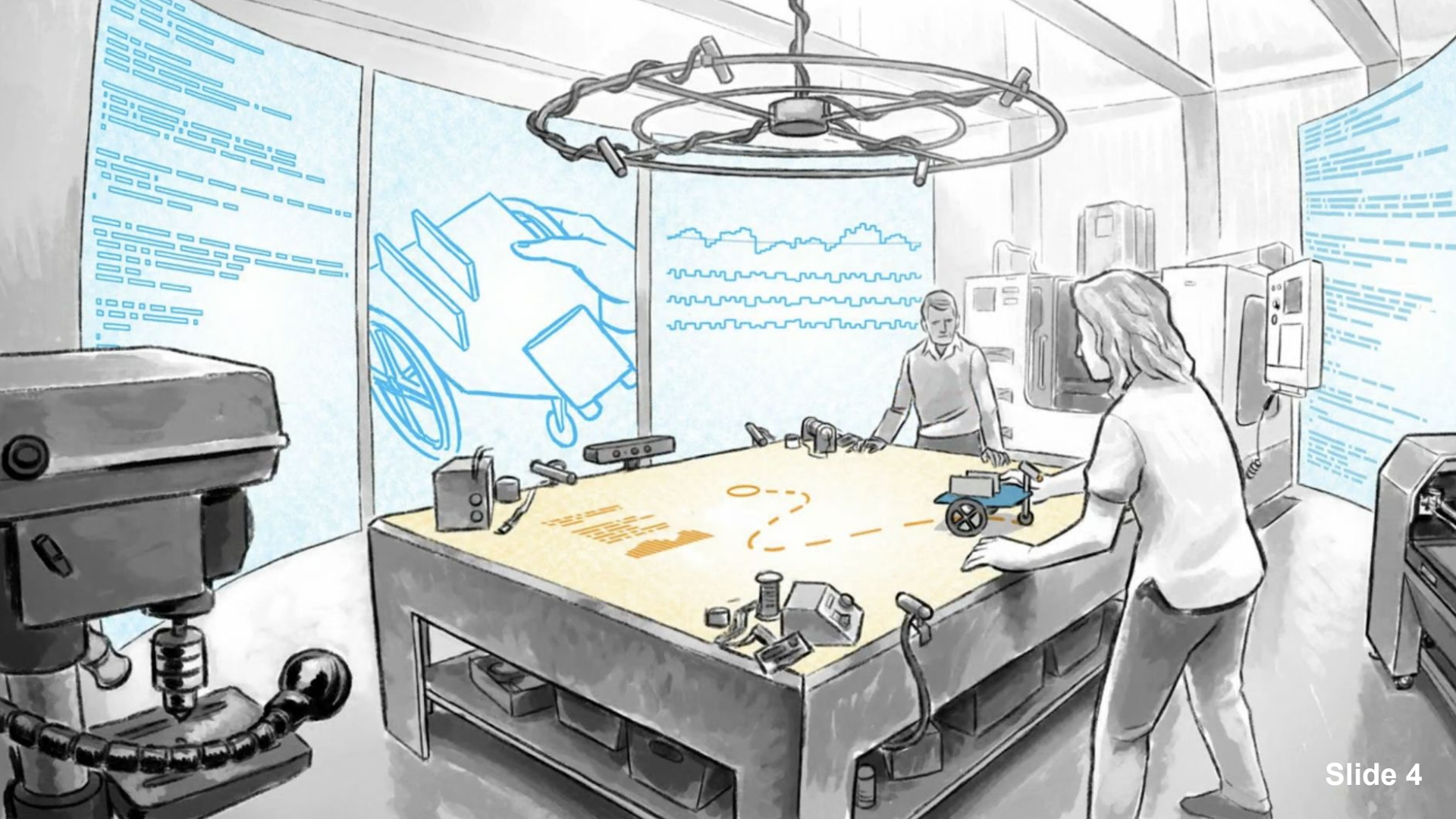
Existing tools have bad visualizations and workflow

Visualizations



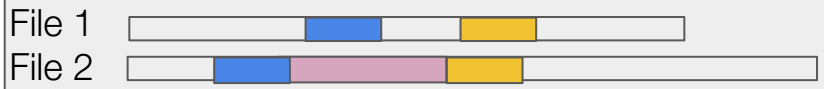
Workflow





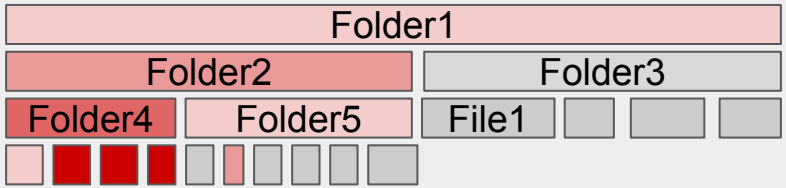
Slide 4

File Graph

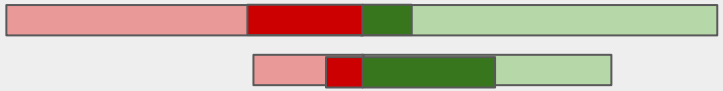


Code/Duplicate View

Hierarchy Graph



Commits Graph



Duplicate Graph

?

View over time

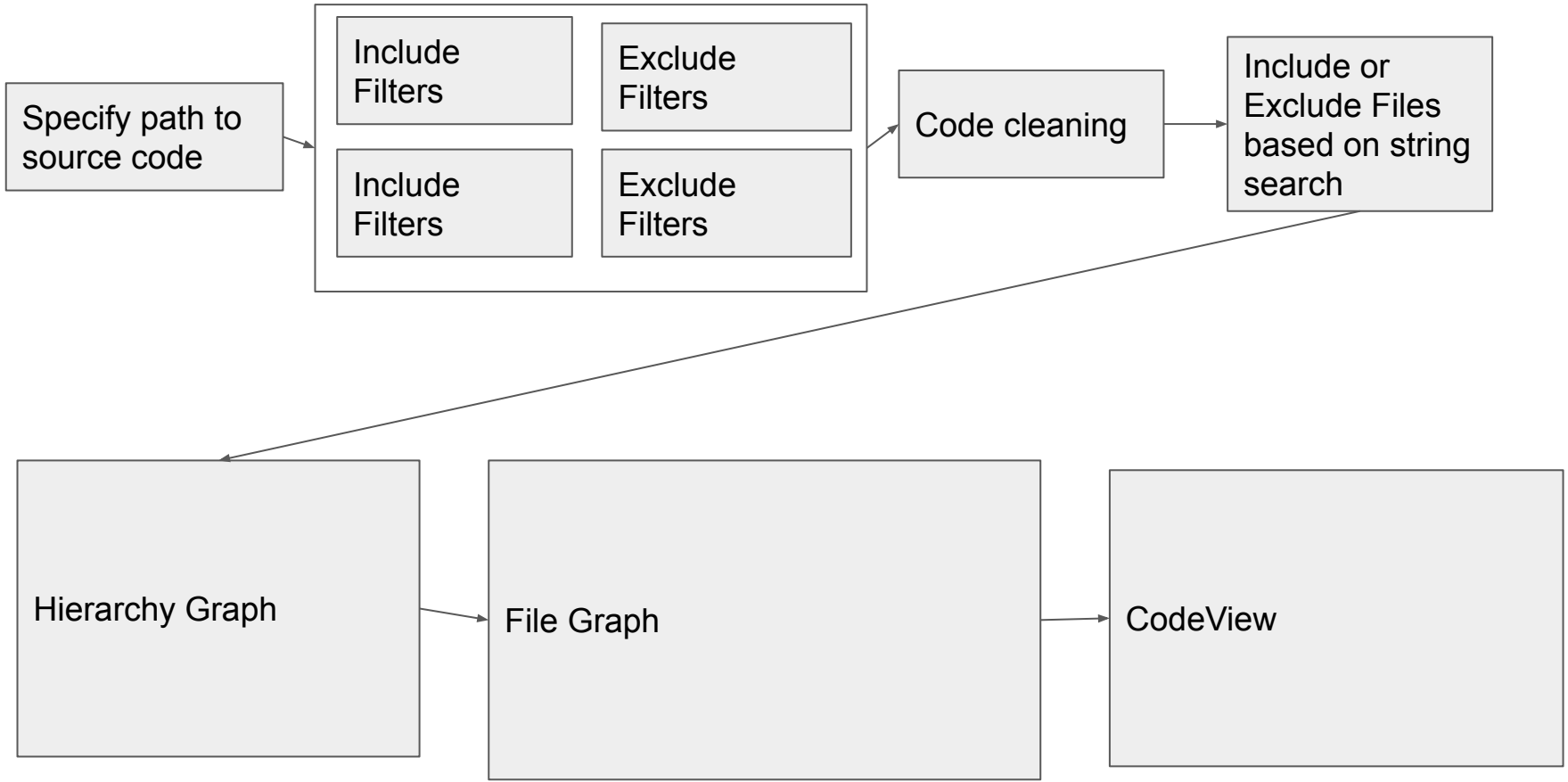
?

Demo

Extra Material

Questions driving the design

Marked in red are questions that probably won't be supported in the tool, but would be nice to have

- **Overview**
  - How much of the application is duplicated?
  - Is it getting better or worse?
  - How does my software system compare to other systems?
- **Where are duplicates located?**
  - Are duplicates concentrated in certain parts of the system?
  - Which parts of the system are duplicated with which?
- **What kind of duplication is there?**
  - What type of duplication is in the files? (e.g., many small fragmented pieces or long chunks)
  - Which files are duplicated with which and what are the duplication patterns between them?
- **What kind of code is duplicated?**
- **Which duplicated code is most problematic?**
  - Which duplicates appear most often?
  - Which duplicates are very long?
- **How has the situation changed over time?**
  - Are specific duplicated sections spreading throughout the codebase? Where? When?
  - Are duplicates getting removed? Where? When?
  - Which parts of the code are getting better or worse?
  - Are inconsistencies being introduced because of duplication?
  - Does duplicated code get cleaned up after a while?
  - Which commits introduce duplication? Which development patterns lead to duplication?

Tool Design - Modular