Trustworthiness in testing and CI/CD

Software Center day Denmark

Kristian Sandahl, Eduard Enoiu, Jean Malm

Projects at Theme 1 Continous Delivery

- #6 Enterprise Scale Continuous Integration and Delivery
- #18 Data Visualization for Continuous Integration
- #29 Modeling and Analyzing Collaborating Machines
- #30 Aspects of Automated Testing

Transparency of the CI/CD flow

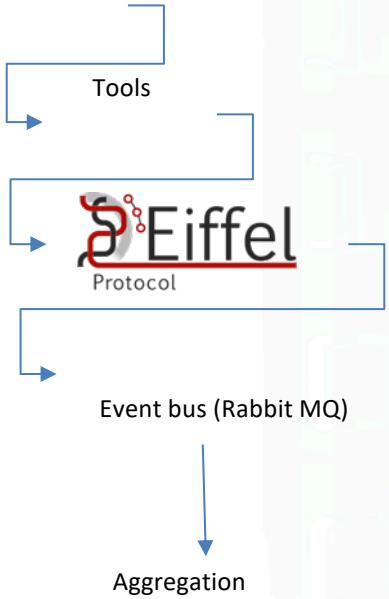
Information needs survey

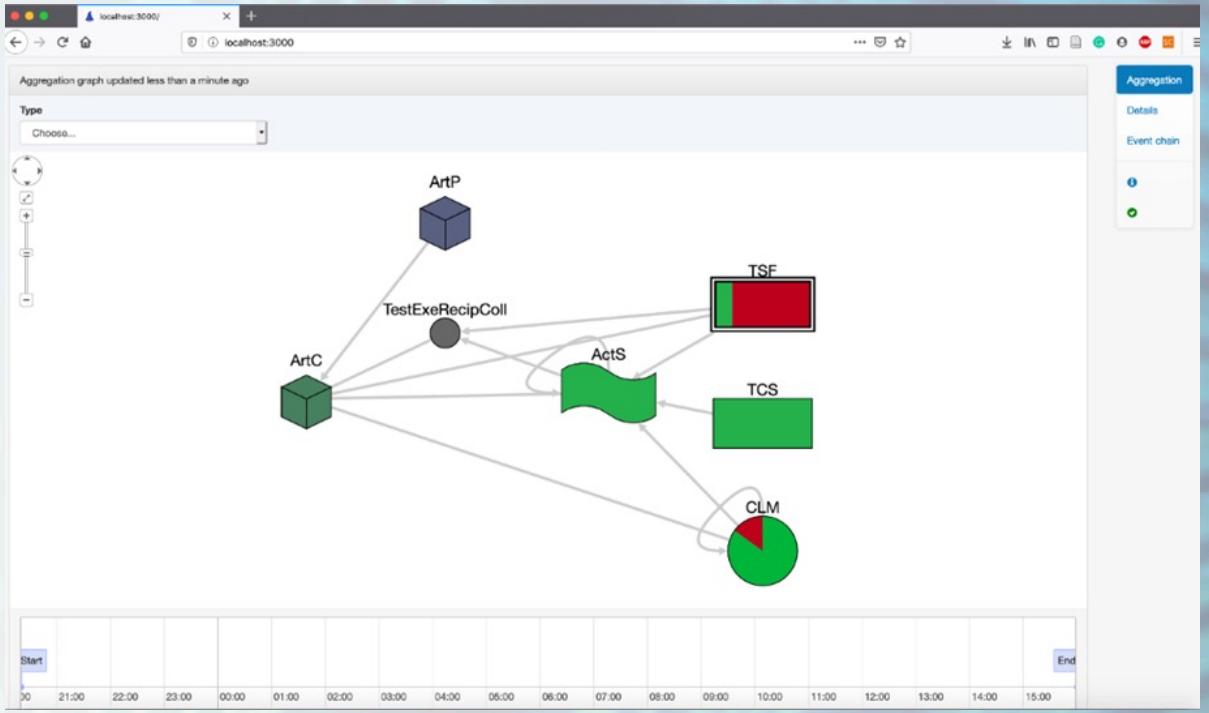
TABLE V

IMPORTANCE, FREQUENCY, EFFORT AND TIME WITH RESPECT TO INFORMATION NEEDS

ID	Information Need	Importance	Frequency	Effort	Time
C1	How much confidence do we have in the release to deploy to the customers?	4,8	**4,9**	**5,0**	15-20
CC6	Is the given feature ready to release to customers?	4,5	**4,6**	**5,0**	15-20
B3	Is the bug fix ready to release to customers?	4,5	**4,6**	**5,0**	15-20
C2	How much confidence do we have in the test suite?	4,1	**4,9**	**5,0**	**>20**
C3	How much confidence do we have in stand-alone projects to be merged into the master branch/baseline?	4,1	**4,7**	**4,8**	**>20**
CC2	What is the status/health of new code changes?	4,1	**4,6**	**4,8**	15-20
CC4	Which change request does the specific commit implements?	4,0	**4,7**	3,5	10-15
CC1	Does the final release to customers include my code?	4,0	3,7	2,5	5-10
T3	In which environment/machine do specific test cases fail?	3,8	**4,7**	**4,5**	>20
T7	Which test cases are flaky?	3,7	**4,7**	**5,0**	**>20**
CC5	Is the given feature implemented?	3,6	**4,6**	**4,5**	10-15
B1	Which bugs have been fixed in the specific release?	3,2	**4,3**	3,3	10-15

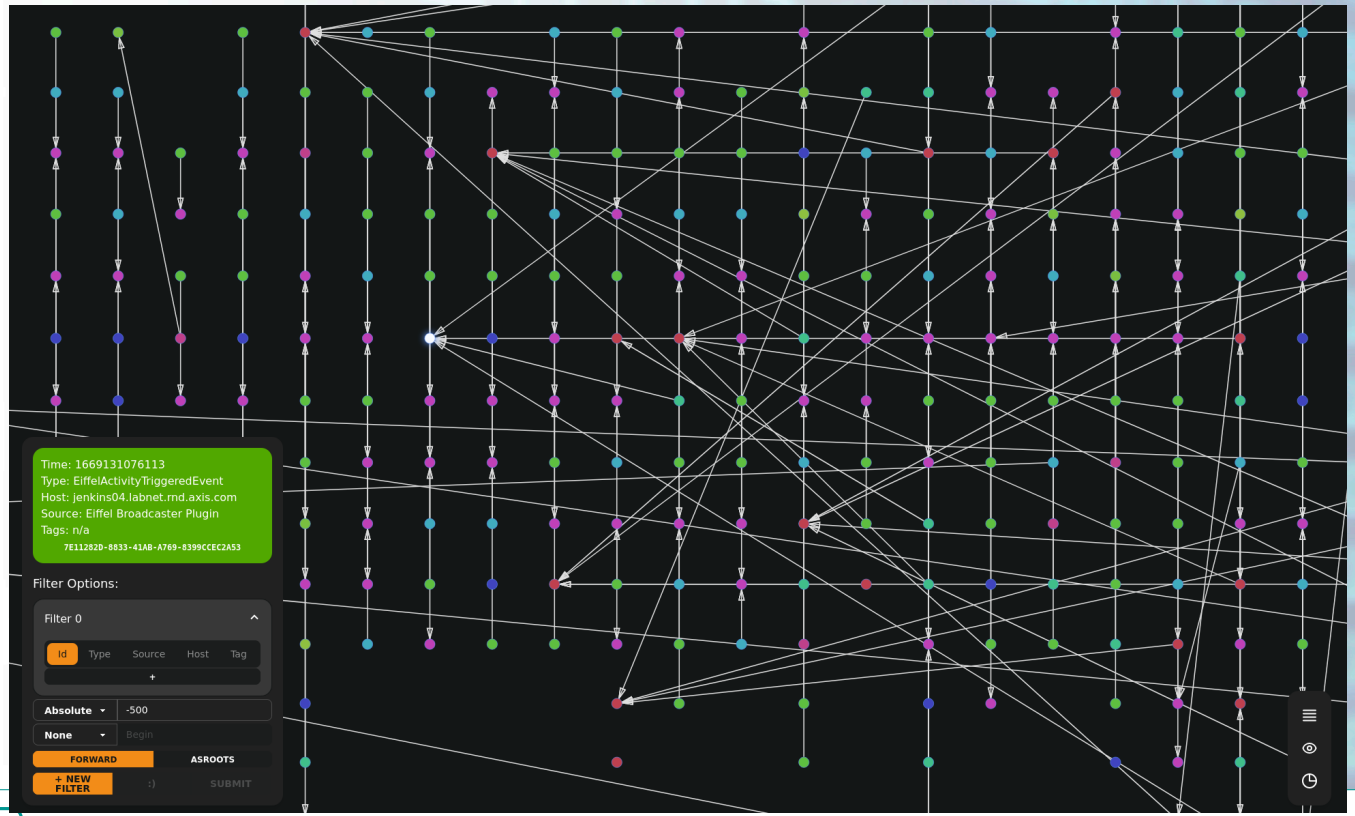
Visualization Eiffel Store





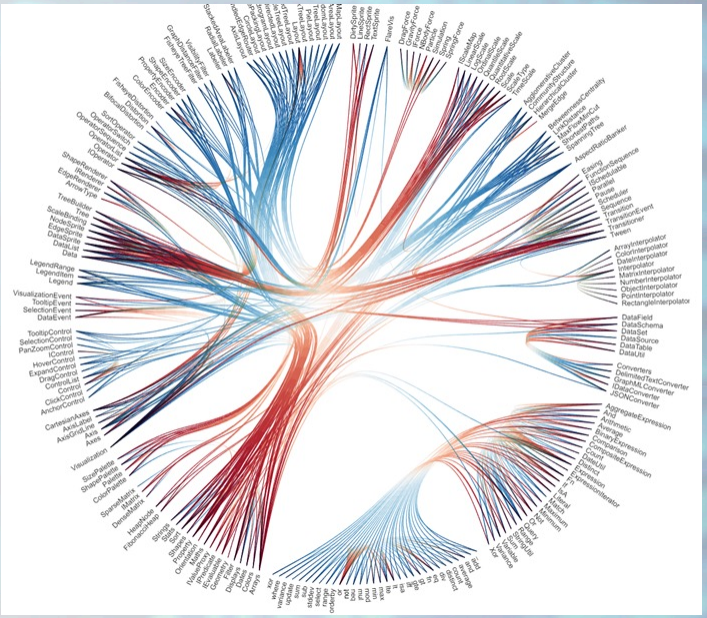
https://github.com/eiffel-community/eiffel-store

Solution by Axis and students from Göteborg



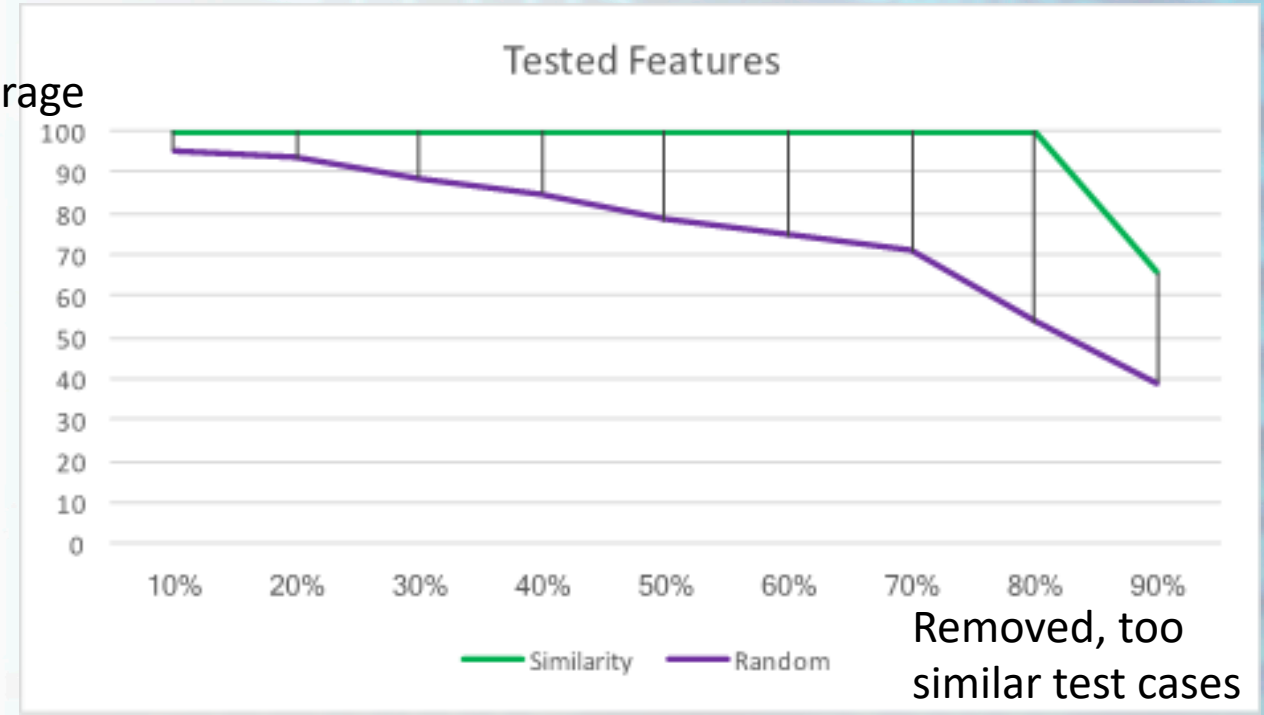
Ongoing and future activities for project #18

- Using fast network visualization methods
- Prediction of Eiffel events



Bottlenecks and trustworthiness in automated testing

Diversity-based testing

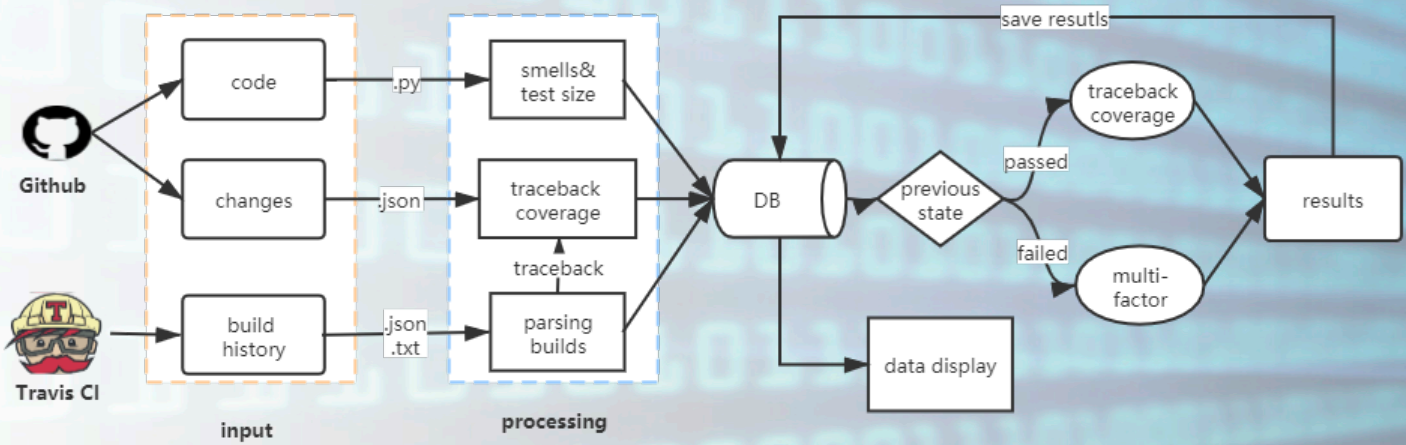


Removed, too similar test cases

https://gitlab.liu.se/azeah70/diversitybasedtesting

Flaky tests detection

Factors:
Traceback coverage
Flip frequency
TC size
Test smells



https://gitlab.liu.se/azeah70/multifactorftdetector

First LiU SWC PhD, 2022-10-06

- Data visualisation in continuous integration and delivery: Information needs, challenges, and recommendations
- Improving Continuous Integration with Similarity-based Test Case Selection
- Empirical Analysis of Practitioners' Perceptions of Test Flakiness Factors
- A Multi-factor Approach for Flaky Test Detection and Automated Root Cause Analysis
- An Evaluation of Machine Learning Methods for Predicting Flaky Tests

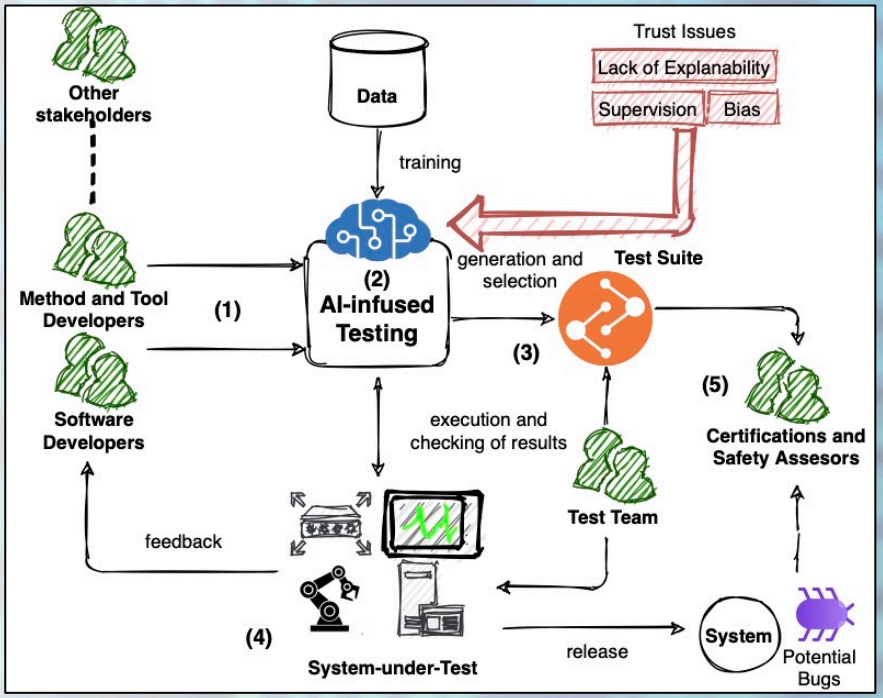
Dr. Azeem Ahmad

Future directions

- Testing of nn modules in a CI context
- How and by whom is confidence in the product built?
- The role of (automated) testing in sustainable SE

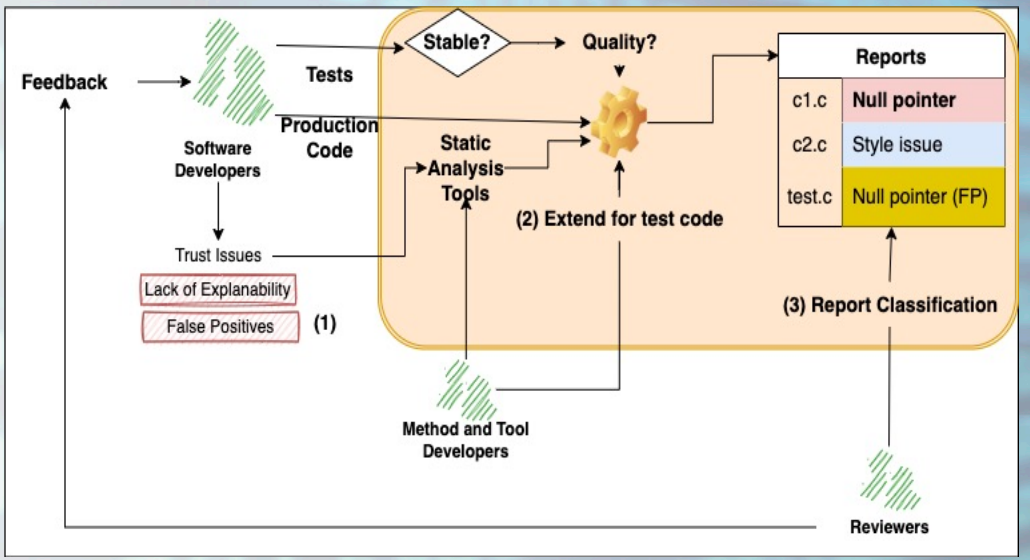
Trustworthy and Human-Centered AI-infused Testing

- Assure that AI-infused testing is sufficiently trustworthy to be deployed at industrial scale
- We identified
  - Challenges: assigning responsibility, bias in decision-making and lack of participation
  - Approaches: explicability, supervision and diversity.
- A checklist for ethical AI in testing and human-centered test design.



Static Analysis of Test Code

- Aim: Leverage Existing Tool Development Effort
  - Goal: Detecting Issues in Test Code Quality.
- We explored existing state-of-the-practice static analysis tools for C/C++
  - Chosen based on usage for production
- We identified:
  - Existing tools' challenges to understand test behaviour (1)
  - General lack of test-specific analysis functionality (2)
- We proposed:
  - Rule-based refutation of noisy reports (3)
  - Enhancements of existing checks (1)



Software Center

Contact: kristian.sandahl@liu.se
jean.malm@mdu.se
eduard.paul.enoiu@mdu.se