

Dissertation summary

Name of candidate

Tomasz Miksa

Name of the supervisor

Andreas Rauber

Title of the dissertation

Verification and Validation of Scientific Workflow Re-executions

Abstract

Scientific experiments performed in the computational science domain require special tooling, software, and workflows that allow researchers to link, transform, visualize and interpret data. Recent studies report that such experiments often cannot be replicated due to differences in the underlying infrastructure. Well-established methodologies for verification and validation known from the software engineering domain are not in use.

We focus on scientific workflows and discuss what needs to be documented and how to create evidence that enables verification and validation of workflow re-executions in a potentially different environment without the necessity of accessing the original environment at the same time.

We devised the VFramework that checks whether the re-executed workflow produces the same result as the original workflow. We automatically create workflow context model that describes workflow and its environment. We verify whether the workflow re-execution was obtained in a compliant way by comparing the context models. We also propose a method for generation of validation requirements that takes into account the data format of the data processed by the workflow and on that basis selects a corresponding set of validation metrics.

We evaluated the VFramework on Taverna workflows from three different domains. We described necessary actions, estimated required effort and demonstrated how these actions can be automated.

The results show that the proposed framework is necessary and can be used for verification and validation of scientific workflow re-executions. Thus the proposed VFramework narrows the gap between the domains of software engineering and scientific workflows.