Web Services for Reproducible Research

Building an Open Source Reproducibility Framework for Document Image Analysis

Marcel Gygli

Reproducible Research has been a research topic within the computational sciences for a long time. It deals with the problem of ensuring that other researchers can independently verify results generated on a computer.

A longstanding problem that reproducible research tries to address is that many published scientific results can not be verified by reading the publication itself. This is due to a set of various reasons: not enough space in the publication to explain all implementation details or aversion to sharing source code publicly. Thus, it becomes ever more difficult for researchers to reproduce results from a presented method, compute results on new data, or use it as part of a new system.

In this thesis, we provide an overview of these issues with a particular focus on the area of Document Image Analysis (DIA), where these problems arise as well. We analyze approaches proposed as solutions to these problems, and identify open issues, especially in the area of providing access to executable versions of research code.

Using this analysis, we design the features of a system that will be able to support DIA researchers in performing reproducible research. This design is implemented in a Web Service framework called DivaServices that allows for the publication and execution of research methods through a unified interface.

The contribution of this thesis is threefold:

First, we analyze how Web Services can be used in the domain of reproducible research and what benefits they offer. This is brought into production by the development of the DivaServices framework as an implementation of such a concept. Second, we design a concept for bringing workflows into the frameworks as they are an essential extension. We show that this concept works by developing a proof-of-concept for the execution of linear workflows. Lastly, we demonstrate the need for an ecosystem of tools and services around such a framework.

Jury:
Prof. Rolf Ingold (thesis supervisor)
Prof. Marcus Liwicki (external co-examiner)
Prof. Bart Lamiroy (internal co-examiner)
Prof. Philippe Cudré-Mauroux (Jury President)