Master

Architektur von Data-Intelligence-Systemen – Integration von Softwarearchitekturmodellen und Data-Science-Workflows

Architecting Data Intelligence Systems – Integrating Software Architecture Models and Data Science Workflows

Motivation — English

Data Intelligence Systems (DIS) are software-intensive systems that use data science workflows to analyze the data in the system. They can change their behavior based on the results of their data analysis. E.g., imagine a software system to identify fraud in an insurance business. That system might learn from labelled data, which customer data correlates with detected fraud. In future requests it can filter potential frauds for proposing human investigation. With more data incoming the system learns to be more precise. Such systems are based on concepts of data science, often specifically artificial intelligence (AI). They are also subject to software engineering in the sense that AI components are developed and operated in the context of—usually complex—software systems.

A major aspect of organizing the structure and behavior of complex software systems, alongside with the organization of their development and operation is the software architecture. When designing a software architecture, goals and requirements of a system are transformed into the major design principles of a software, usually breaking it down into interconnected components.

The goal of this thesis is to develop a method to integrate software architecture descriptions in the Unified Modeling Language (UML) with data science workflows using RapidMiner Studio.

In this thesis you will investigate the following research questions:

1. How can the interfaces of operators and workflows in data science workflows be described for the use in software architectures?
2. How can these descriptions be integrated with software architecture models wrt. structure and behavior modeling?
3. How can the execution of systems modeled with this approach be monitored?

Knowledge required to carry out the work: Modeling with UML, Java
Helpful knowledge: Data Science, Software Modeling, RapidMiner Studio, UML Component Models, UML State Charts, UML Activity Diagrams

Organization

Contact:
Dr. Marco Konersmann <konersmann@uni-koblenz.de>

Literatur

[1] RapidMiner Studio, `https://rapidminer.com/products/studio/`

[2] Eclipse Papyrus Modeling Environment, `https://www.eclipse.org/papyrus/`

[3] OMG Unified Modeling Language Specification Version 2.5.1, `https://www.omg.org/spec/UML/2.5.1/PDF`