
Bachelor/Master Thesis Topic

Analyzing the Language of Data Analysis Workflows

Motivation and Background

The concept of Data Analysis Workflows (DAW) is well known [1] and different real world implementations exist. The principle idea of a DAW is to provide a solution strategy to a complex data analysis problem (e.g., identifying flooding areas based on satellite pictures) by using already established programs to solve sub problems of the overall problem (e.g. removing pictures of clouds from the analysis, finding water, comparing other pictures, ...). Therefore, different languages to express such DAWs exist and most of the languages also provide tooling to support the execution of the DAW. A generalization of such languages is partly present [2].

Goals

The aim of the thesis is to provide a detailed analysis of current DAW languages by providing a mapping to the Meta Model (generalization) in [2]. Also, an extension of the Meta Model (EMF Ecore) is needed. Possible extensions are: A concept and/or implementation of an automatic translation from one language to another. An analysis of the powerfulness of the different languages. A identification of limitations of the analyzed languages.

Description of the Task

The specific tasks are:

- Analysis of the Meta Model
- Description of the DAW languages based on available literature.
- Mapping of the languages.

Research Type

Theoretical Aspects:	****
Industrial Relevance:	*****
Implementation	**

Prerequisite

The student should be enrolled in the bachelor of computer science program, and has completed the required course modules to start a bachelor thesis.

Skills required

Literature analysis, very careful documentation of results, working carefully with citations, interest in languages in computer science.

Contacts

Marcus Hilbrich (marcus.hilbrich@informatik.hu-berlin.de)

Software Engineering Group, Institut für Informatik, Humboldt-Universität zu Berlin

References

[1] Leser, Ulf; Hilbrich, Marcus; Draxl, Claudia; Eisert, Peter; Grunske, Lars; Hostert, Patrick; Kainmüller, Dagmar; Kao, Odej; Kehr, Birte; Kehrer, Timo; Koch, Christoph; Markl, Volker; Meyerhenke, Henning; Rabl, Tilmann; Reinefeld, Alexander; Reinert, Knut; Ritter, Kerstin; Scheuermann, Björn; Schintke, Florian; Schweikardt, Nicole; Weidlich, Matthias
The Collaborative Research Center FONDA Journal Article
In: Datenbank-Spektrum, no. 1610-1995, 2021.

[2] Hilbrich, Marcus; Müller, Sebastian; Kulagina, Svetlana; Lazik, Christopher; Mecquenem., Ninon De; Grunske, Lars
A consolidated View on Specification Languages for Data Analysis Workflows Proceeding Forthcoming
Automated Software Re-Engineering (ISoLA2022 · ASRE) (accepted), Forthcoming.