

Gemeinschaftliche Modellierung von Offene-Daten-Pipelines - JValue-PML

Problemstellung

Offene Daten kommen in den unterschiedlichsten Formaten daher und sind oft schlecht definiert. Vor ihre Verwertung ist ein arbeits- und kostenintensiver Daten-Engineering-Prozess geschaltet. Statt jede Nutzer:in den Daten-Engineering-Prozess individuell durchführen zu lassen, wäre es besser, dies gemeinschaftlich in einem Open-Source-Projekt zu tun. Dazu fehlt aber eine Modellierungssprache mit Ausführungsumgebung für Daten-Pipelines.

[hier Bild einfügen, Bild wird vom BMDV
Eingefügt, Format: 1200 x 675 Pixel]

Bildunterschrift: Arial 7 (in Klammern: Quelle:)

Projektziel

Wir wollen eine Modellierungssprache für Daten-Pipelines entwickeln, mit deren Hilfe Pipeline-Modelle erstellt werden können, sowie einen Compiler und Laufzeitumgebung, mit deren Hilfe die Pipelines ausgeführt werden können. Die Modellierungssprache muss eine textuelle Syntax haben, so dass interessierte Parteien über etablierte Open-Source-Infrastruktur gemeinschaftlich an Pipeline-Modellen arbeiten können.

Durchführung

Dazu werden wir (1) eine Modellierungssprache definieren, (2) einen Compiler entwickeln, und ein (3) Laufzeitsystems entwickeln, welches vom Compiler instrumentiert wird, um eine ausführbare Pipeline zu erhalten. Wir werden (4) mittels studentischer Projekte zu Mobilitätsdaten demonstrieren, dass die Pipeline-Modelle auf Basis unserer Sprache gemeinschaftlich entwickelt werden können.

Verbundkoordinator

Professur für Open-Source-Software,
Friedrich-Alexander-Universität Erlangen-Nürnberg

Projektvolumen

Summe [wird vom BMDV eingetragen]
(Förderquote wird vom BMDV eingetragen)

Projektlaufzeit

11/2022 – 10/2023

Projektpartner

Professur für Open-Source-Software,
Friedrich-Alexander-Universität Erlangen-Nürnberg

Ansprechpartner

Professur für Open-Source-Software
Prof. Dr. Dirk Riehle
Telefon: +49 9131 85 28390
E-Mail: dirk.riehle@fau.de