

Entwicklung einer Graphdatenbank zur Speicherung von Prov-Graphen

Hintergrund Die Nachvollziehbarkeit und Reproduzierbarkeit von Forschungsergebnissen sind zentrale Qualitätskriterien in allen Bereichen der datengetriebenen Wissenschaft. Das Tool-Set **provQCD** wurde speziell für die Lattice-QCD-Community entwickelt, um automatisch **Provenance**-Informationen zu erfassen, in ein **W3C PROV**-konformes Format zu übersetzen, zu bündeln und schließlich grafisch darzustellen. Grundlage hierfür sind sogenannte **ProvCards**. Zur automatischen Erhebung und Auswertung von Provenance-Daten haben wir das Konzept der **ProvDecks** entwickelt. Dieses Konzept basiert auf drei miteinander verbundenen Strukturen:

- **ProvCards**, die als maschinenlesbare Metadateneinheiten einzelne Entitäten, Aktivitäten oder Agenten beschreiben;
- **ProvDecks**, die Sammlungen von Provenance-Informationen (inklusive ProvCards, Metadaten und Graphstrukturen) für Versionierung und Austausch bereitstellen;
- **ProvGraphs**, die automatisch erzeugte Visualisierungen der Abhängigkeiten zwischen Daten, Prozessen und Akteuren ermöglichen.

Aufgabenstellung **provQCD** erzeugt Provenance-Informationen in Form von **ProvCards** und bündelt diese zu **ProvDecks**. Derzeit werden die Graphen als **.dot**-Dateien erstellt und als **pdf** gespeichert. Für moderne, datenintensive Forschungsinfrastrukturen wird jedoch häufig eine Graphdatenbank eingesetzt, um große, vernetzte Daten effizient zu speichern, abzufragen und zu visualisieren. Die zentrale Forschungsfrage lautet daher: *Wie lässt sich der in **provQCD** erstellte Provenance-Graph in einer Graphdatenbank modellieren, abfragen und in bestehende Provenance-Workflows integrieren?*

Ziel der Arbeit ist die Konzeption, Implementierung und Evaluation einer Graphdatenbank-Komponente innerhalb der **provQCD**-Toolbox, die ProvGraphs in einem standardkonformen Format abbildet und zugleich die bestehenden Module erweitert.

Formalia

- Ansprechpartner:
 - Tanja Auge (Fakultät für Informatik und Data Science, tanja.auge@ur.de)
 - Christian Kindler (Fakultät für Physik, christian.kindler@ur.de)
- Voraussetzungen:
 - Interesse an Provenance-Konzepten für reale Szenarien
 - Fähigkeit zur eigenständigen Recherche und Auswertung wissenschaftlicher Literatur
 - Erste Erfahrungen mit **W3C PROV/ILDG** (oder Bereitschaft zur vertieften Einarbeitung)
 - Gute Python-Kenntnisse
 - Erfahrung mit graphenbasierten Visualisierungstools (oder Bereitschaft zur vertieften Einarbeitung)

Literatur

- Auge et al.: Provenance for Lattice QCD Workflows. <https://dl.acm.org/doi/10.1145/3543873.3587559>
- Auge et al.: Provenance for Lattice QCD Workflows – An Update. <https://dl.acm.org/doi/10.1145/3736229.3736268>
- Kleinstüber et al.: Managing Provenance Data in Knowledge Graph Management Platforms. <https://link.springer.com/article/10.1007/s13222-023-00463-0>
- Kleinstüber et al.: A Provenance Management Framework for Knowledge Graph Generation in a Web Portal. <https://dl.gi.de/items/22c15a52-aca2-4dc9-957b-86cf115d5b97>