

Erweiterungen der provQCD-Toolbox

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**. **provQCD** besteht aktuell aus den Modulen:

- **metadata_extractor**: Extrahiert Provenance-relevante Metadaten aus Log-/Metadata-Dateien.
- **translator**: Wandelt Rohdaten in W3C-PROV-Terminologie und ProvCards um.
- **merger**: Bündelt ProvCards zu ProvDecks.
- **registrar**: Soll ProvCards in einer durchsuchbaren Datenbank speichern und als Schnittstelle zu weiteren Datenbanken dienen.
- **graph_generator**: Erzeugt eine .dot-Datei des Provenance-Graphen.
- **visualizer**: Visualisiert den Provenance-Graphen als pdf.

Ziel dieser Bachelorarbeit ist die Weiterentwicklung der **graph_generator**- und **visualizer**-Module sowie die Implementierung zusätzlicher Datenbank-Komponenten zur Verwaltung der Agenten und ProvCards zur Entwicklung einer allgemeinen Provenance-Pipeline.

Aufgabenstellung Die Arbeit soll die folgenden Kernfunktionen realisieren:

1. **Agenten-Datenbank**: Aufbau einer relationalen Datenbank (z.B. PostgreSQL/SQLite) zur Verwaltung von Agenten (Benutzer, Software, Hardware). Das Modell muss Agenten-IDs, Rollen, Zugehörigkeiten sowie weitere Metadaten abbilden.
2. **ProvCard-Datenbank**: Erweiterung der **registrar**-Komponente zur Speicherung der ProvCard Metadaten inklusive der Zuordnung der Speicherpfade pro ID.
3. **ProvGraph in PROV-JSON**: Erweiterung des **graph_generator** auf einen vollständig konformen PROV-JSON Output.
4. Interaktive Visualisierungen (Zoom, Filter, Layout-Auswahl, SVG-Export) für die graphische Darstellung.

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>