

Bachelorarbeit

Hands-on mit Linked Data: Modellieren mit Protégé OWL und neo4j

Kurzbeschreibung

In dieser Arbeit kommen zwei weit verbreitete Anwendungen (neo4j und OWL Protégé) zur Modellierung von komplexen und vernetzten Informationen zum Einsatz. Nach dem Erlernen der jeweiligen Grundlagen durch Arbeit mit Tutorials, sollen Gemeinsamkeiten und Unterschiede der Anwendungen durch Übertragung der Tutorial-Modelle in die jeweils andere Anwendung herausgearbeitet werden und anschließend ein weiterer vorgegebener Datensatz in beiden Anwendungen modelliert werden. Inhalt der schriftlichen Ausarbeitung ist eine Beschreibung der Gemeinsamkeiten und Unterschiede der jeweiligen Modellierprozesse und resultierenden Modelle, des jeweiligen Komforts sowie der dadurch möglichen Analysemöglichkeiten und -prozesse.

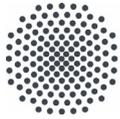
Hintergrund und Details

Nicht nur im Bereich Industrie 4.0 sowie beim Austausch von Informationen zwischen Maschinen und entlang komplexer Liefer- und Wertschöpfungsketten gewinnen interoperable Datenformate an Bedeutung: Die Möglichkeit große und komplexe Wissensbestände und Datensammlungen computerunterstützt eindeutig auswerten zu können, gewinnt im Zuge der Digitalisierung in einer Vielzahl von Kontexten an Bedeutung. Dafür müssen die entsprechenden Informationselemente in einem maschinell eindeutig interpretierbaren Format vorliegen. Als von individuellen technischen Realisierungen unabhängige formale Wissensrepräsentation und als Austauschformat zwischen Akteur*innen und Anwendungen setzt sich zunehmend die Modellierung des Wissens in Form von Linked Data durch. Sogenannte Ontologien stellen dabei sowohl die konzeptionelle Modellierungsumgebung dar, als auch fertig modellierte Sachverhalte und Wissensbestände.

Der Ontologie Editor Protégé OWL hat seine Ursprünge in der Biomedizin, ist inzwischen aber vor allem im akademischen Bereich in fast alle Disziplinen vorgedrungen, um komplexe und häufig hochvernetzte Informationen zu modellieren und computerunterstütztes Reasoning (Ableitungen, Inferenz von Folgerungen) über den Daten zu betreiben.

Die Graph Datenbank neo4j hat erst kürzlich wieder in einer Finanzierungsrunde mit 325 Millionen US-Dollar laut eigenen Angaben die größte Investitions-Transaktion zugunsten eines privaten Datenbank Unternehmen jemals erzielen können. Unter anderem erhielt neo4j viel Aufmerksamkeit im Zusammenhang mit den Panamapapers, als ein Konsortium von 370 Reporter*innen mit Hilfe von neo4j Information aus 11,5 Millionen geleakten Dokumenten extrahierte und in einen Graphen bestehend aus 840.000 Knoten und 1,3 Millionen Kanten (Relationen) zusammenführte. Dadurch wurden die Informationen besser analysierbar, Netzwerke und Off-Shore Steueroasen wurden sichtbar und die Erkenntnisse der investigativen Journalist*innen wurden mit dem prestigeträchtigen Pulitzer-Preis gewürdigt.

Beide oben genannte Anwendungen unterstützen Linked Data konforme Modellierung von Sachverhalten, aber es liegen ihnen jeweils auf den ersten Blick sehr unterschiedliche Konzepte und Terminologie zugrunde.



Ziel der Arbeit ist zunächst die Erlernung der Grundlagen beider Anwendungen mit Hilfe von zur Verfügung gestellten Tutorials. Danach sollen die aus den Tutorials resultierenden Modelle in der jeweils anderen Anwendung erstellt werden. Dabei werden wesentliche unterschiedliche Modellier-Paradigmen und -Konzepte deutlich, so dass die gefundenen Unterschiede im Modellierungsprozess und den resultierenden Modellen beschrieben werden können. Als letztes sollen die Studierenden in CSV (Comma Separated Values, Excel) vorliegende Daten in beide Anwendungen importieren, diese mit weiteren gegebenen Linked Data Vokabularen verbinden und auch hier eine vergleichende Beschreibung der Prozesse, des Komforts und der Analysemöglichkeiten der resultierenden Modelle verfassen.

Aufgabenstellung

- Nutzung mindestens der beiden Software-Suiten „Protégé OWL“ und „neo4j“, ggfs. auf Wunsch noch weitere Alternativen (z.B. FluentEditor o.a.).
- Abbildung/Nachbauen der berühmten Pizza-Ontologie (Protégé-Tutorial) und des Actor-Movie Graphen (neo4j-Tutorial) in allen verwendeten Software-Suiten.
- Vergleichende Beschreibung der jeweiligen Modellierprozesse und der resultierenden Modelle.
- Danach: Import vorliegender Excel-Daten (.csv) in die jeweiligen Suiten und Anbindung an gegebene externe Linked Data Vokabulare und vergleichende Beschreibung der Prozesse, des Komforts und der Analysemöglichkeiten der resultierenden Modelle.

Profil der Arbeit

Themenart: theor./exp.

Voraussetzungen

Genügend Englischkenntnisse um den Tutorials folgen zu können:

<https://neo4j.com/developer/cypher/guide-cypher-basics/>, und

<https://www.michaeldebellis.com/post/new-protege-pizza-tutorial>

Hinweise

Die Arbeit kann auf Deutsch oder Englisch verfasst werden.

Betreuung



M.A. Dipl.-Ing. (FH) Jan Wunderlich

jan.wunderlich@ids.uni-stuttgart.de

Tel.: 0711 – 685 60707

www.ids.uni-stuttgart.de