



FernUniversität in Hagen

– Fakultät für Mathematik und Informatik –

Schema für Graph-Datenbanken

Seminar „Discovering Big Data“

Lehrgebiet Datenbanken und Informationssysteme

vorgelegt von

Sarah Kreutzke

Matrikelnummer: 4345339

Betreuung : Prof. Dr. Uta Störl

EXPOSÉ

1.1 FORSCHUNGSTHEMA

Graph-Datenbanken unterscheiden sich von relationalen Datenbanken dadurch, dass sie nicht auf ein fest definiertes Schema angewiesen sind. Es kann jedoch sinnvoll sein, ein Schema für Graph-Datenbanken zu erstellen, um die Struktur und Optimierung der Datenbank zu verbessern. Bisher gibt es jedoch keine einheitliche Methode oder standardisierte Sprache für das Schema von Graph-Datenbanken. Dies kann zu Herausforderungen führen, wenn es darum geht, die Datenbank effizient zu verwalten und Abfragen durchzuführen. Es ist daher wichtig, dass Unternehmen, die Graph-Datenbanken verwenden, sorgfältig überlegen, wie sie ihr Schema gestalten, um die Leistung und Skalierbarkeit ihrer Datenbank zu optimieren. Durch die Definition eines klaren Schemas können potenzielle Probleme vermieden und die Effizienz der Datenbank verbessert werden.

1.2 ZIELSETZUNG

Das Ziel dieser Studie besteht darin, die Gemeinsamkeiten und Unterschiede der verschiedenen Ansätze für Graph-Datenbank-Schemata aufzuzeigen und zu untersuchen, ob es eine Grundstruktur gibt, auf die alle Ansätze aufbauen. Dafür wird eine Literaturrecherche durchgeführt.

1.3 FORSCHUNGSSTAND UND THEORETISCHE GRUNDLAGEN

Graph-Datenbanken sind darauf spezialisiert, vernetzte Daten zu verarbeiten und bieten eine einfache Möglichkeit, Beziehungen zwischen verschiedenen Datentypen zu modellieren. Die Hauptbestandteile einer Graph-Datenbank sind Knoten und Kanten, wobei Kanten in der Regel gerichtet sind. Diese Datenbanken verfügen über zahlreiche Traversierungsalgorithmen, die bei der Mustererkennung hilfreich sind. Ein bedeutender Vorteil von Graph-Datenbanken liegt in ihrer Flexibilität, die durch das Fehlen eines festen Schemas ermöglicht wird. Allerdings kann dies auch Nachteile mit sich bringen, da es schwierig sein kann, Integritätsbeschränkungen durchzusetzen. Es besteht die Möglichkeit, dass Knoten in der Datenbank mit beliebigen anderen Knoten verbunden werden können, unabhängig von ihrer Sinnhaftigkeit [RH+17].

Es gibt unterschiedliche Herangehensweisen und Bemühungen, ein Schema für Graph-Datenbanken zu definieren: Roy-Hubara *et al.* haben ein Schema auf Basis eines Entity-Relationship-Modells definiert [RH+17]. Das World Wide Web Consortium hat im Jahr 1998 ein Schema für eine Art der Graph-

Datenbank, dem Resource Description Framework (RDF) festgelegt. Dieses Schema wird kontinuierlich überarbeitet, wobei die aktuellste Version aus dem Jahr 2014 stammt [BG]. Für Property-Graph-Datenbanken gibt es keinen einheitlichen Ansatz zur Schema-Definition. Eine formale Definition für diese Art von Graph-Datenbank wurde beispielsweise von Renzo Angles im Jahr 2018 präsentiert [Ang].

1.4 KONZEPT

Basierend auf einer Literaturrecherche wird in dieser Studie eine Untersuchung des aktuellen Forschungsstandes zum Thema „Schema für Graph-Datenbanken“ durchgeführt. Dabei werden verschiedene Methoden zur Erstellung eines Graph-Datenbank-Schemas vorgestellt und miteinander verglichen. Um eine umfassende Darstellung zu gewährleisten, ist es wichtig, sowohl allgemeine Literatur über das Schema von Graph-Datenbanken als auch spezifische Literatur über RDF und Property-Graph-Datenbanken zu berücksichtigen.

1.5 ZEITPLAN

Nach der Themenvergabe Ende März wurde die Literaturrecherche Mitte April abgeschlossen. Anschließend wurde eine vorläufige Gliederung und das Exposé erstellt.

Nach der Abgabe des Exposé Ende April wird mit der eigentlichen Hausarbeit begonnen. Ziel ist es die Arbeit bis Mitte Juni abzuschließen, um vor der finalen Abgabe zwei Wochen für letzte Korrekturen und kleine Änderungen zu haben.

Zeitplan Hausarbeit „Schema für Graph-Datenbanken“	März		April				Mai				Juni				Juli	
	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	
Themenauswahl und -vergabe																
Literatursuche																
Literatursichtung																
Literaturrecherche																
Vorläufige Gliederung planen																
Exposé schreiben und einreichen																
Gliederung erstellen																
Einleitung verfassen																
Hauptteil mit Unterkapiteln verfassen																
Abgabe vorläufige Ausarbeitung																
Schluss schreiben																
Korrektur lesen																
Formatierungen prüfen																
Quellenverzeichnis und Literaturangaben prüfen																
Abgabe finale Ausarbeitung																

Abbildung 1.1: Zeitplan der Hausarbeit „Schema für Graph-Datenbanken“

LITERATUR

- [Ang] Renzo Angles. "The Property Graph Database Model". In: (). URL: <https://www.ceur-ws.org/Vol-2100/paper26.pdf>.
- [BG] Dan Brickley und R. V. Guha. *RDF Schema 1.1*. URL: <https://www.w3.org/TR/rdf11-schema/>.
- [RH+17] Noa Roy-Hubara, Lior Rokach, Bracha Shapira und Peretz Shoval. "Modeling Graph Database Schema". In: *IT Professional* 19.6 (Nov. 2017), S. 34–43. ISSN: 1520-9202. DOI: 10.1109/MITP.2017.4241458. URL: <http://ieeexplore.ieee.org/document/8123463/> (besucht am 20.04.2024).