

Evaluation von Graphdatenbanken und MySQL am Beispiel von Kookkurrenzgraphen

Martin Junghanns

`martin.junghanns@studserv.uni-leipzig.de`

Sascha Ludwig

`s.ludwig@studserv.uni-leipzig.de`

Robert Schulze

`robert.schulze@studserv.uni-leipzig.de`

13. Januar 2012

1 Ziel

Ziel des Praktikums ist die Evaluation verschiedener Graphdatenbanken und eine Gegenüberstellung mit relationalen Datenbanken. Als Vertreter der Graphdatenbanken werden Neo4j¹, Dex² und OrientDB³ näher betrachtet, MySQL⁴ wird als relationale Datenbank in den Vergleich einbezogen.

Konkrete Ziele des Praktikums sind das Kennenlernen der verschiedenen Graphdatenbanken, der zu Grunde liegenden Datenmodelle und der angebotenen Abfragetechniken. Für den experimentellen Vergleich sollen verschiedene Kookkurrenzgraphen des Wortschatz Leipzig Projektes in die Datenbanken importiert und mit den angebotenen Techniken abgefragt werden. Dabei werden ausgewählte SQL Anfragen in API Calls, aber auch in deklarative Anfragesprachen des jeweiligen Herstellers umformuliert und hinsichtlich ihrer Ausführungszeit verglichen.

Wir vermuten, dass Graphdatenbanken mit zunehmendem Umfang der Graphen und insbesondere bei rekursiven Anfragen durch die Vermeidung von Verbundoperationen besser skalieren als MySQL.

¹<http://neo4j.org/>

²<http://www.sparsity-technologies.com/dex>

³<http://www.orienttechnologies.com/orient-db.htm>

⁴<http://www.mysql.de/>

2 Zeitplan

bis	Aufgabe
31.10	Spezifizieren der genauen Aufgabenstellung, Treffen mit D. Goldhahn
31.10	Spezifizieren, welche Queries an das DBMS gestellt werden sollen
31.10	Recherche verschiedener Graphdatenbanken
07.11	Erstellen eines Code-Repositories (github) für kollaboratives Arbeiten
14.11	Einarbeitung in die jeweiligen Java APIs
21.11	Erstellen einer Klassenstruktur für das gesamte Projekt
21.11	Implementieren der Importer von MySQL in die jeweiligen GraphDBs
05.12	Implementieren der Queries in der native API der jeweiligen GraphDBs
12.12	Implementieren der Queries in eventuell vorhandenen High-Level Sprachen
19.12	Implementieren von automatisch durchführbaren Benchmarks mit anschließender automatischer Generierung von Performance-Plots
09.01	Optimierung für jedes GraphDBs
09.01	Refaktorisierung des gesamten Projektes
10.01	Erstellen und Überarbeiten der Java-Code-Dokumentation
13.01	Durchführen der Benchmarks und Erstellen der Plots
16.01	Erstellen des Abschlussberichtes
16.01	Erstellen der Abschlusspräsentation
20.01	Halten der Abschlusspräsentation