

Linked Data Performance

Fachgebiet: Informatik
Betreuer: Dr. Urs Sauter
Experte: Prof. Dr. Spichiger Andreas

Die Aufgabe der Bachelorthesis «Linked Data Performance» war es, ein Testsystem auf einem virtuellen Server zu installieren. Dabei wurden zwei RDF-Plattformen und ein herkömmliches Datenbanksystem eingerichtet. Die Messresultate werden in einer relationalen Datenbank gespeichert und nach den Kriterien Bandbreite der Netzwerkverbindung, Leistungsfähigkeit des Clients und Browser ausgewertet.

Auf dem virtuellen Server `vmnabu.bfh.ch` wurde das Testsystem eingerichtet. Dazu haben wir ein bestehendes Datenset eingesetzt, welches von `data.gemeindezukunft.ch` zur Verfügung gestellt wird. Dieses Datenset haben wir in den Fuseki Triplestore eingebunden. Zudem ist eine relationale Abbildung davon in einer MS SQL Datenbank gespeichert. Die installierte D2RQ-Plattform kann die relationale Datenbank in einen RDF-Graphen mappen. Um einen Vergleich mit relationalen Abfragen herzustellen, wurde zudem eine kleine Serverapplikation mit Node.js erstellt.

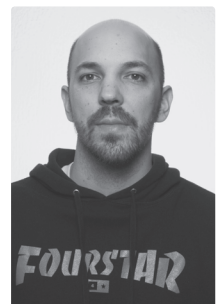
Da die Datenbanken über verschiedene Browser abgefragt werden sollen, haben wir uns entschieden, ein clientseitiges Testtool zu entwickeln. Das Tool kann die Abfragen an die verschiedenen Plattformen senden und deren Zeitdauer messen. Die Metadaten zu einer Abfrage können im Tool ausgewählt werden.

Wir haben vier Abfragen mit unterschiedlicher Komplexität und Datenmenge ausgearbeitet. Die Messdaten wurden innert drei Tagen aufgezeichnet. Dabei haben wir Messungen am Netzwerkanschluss der BFH in Bern und von privaten Anschlüssen aus durchgeführt.

Die Messdaten wurden mit Hilfe von SQL Abfragen und Microsoft Excel ausgewertet. Wir haben festgestellt, dass die RDF Plattformen D2RQ und Fuseki noch nicht mit einer relationalen Datenbank mithalten können. Insbesondere bei Anfragen mit grosser Komplexität, sind die RDF Plattformen um ein Vielfaches langsamer. Die Messungen am BFH Netzwerkanschluss waren schneller als Messungen, welche von Zuhause aus gemacht wurden. Bei den Browsern konnte kein Favorit festgestellt werden. Wie erwartet, liefert ein Client mit mehr Rechenleistung auch durchschnittlich schnellere Resultate.



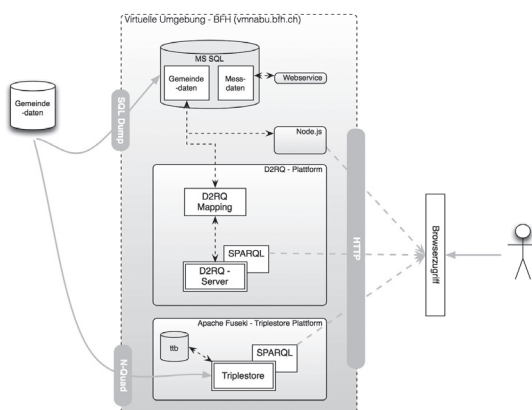
Sahar Faruk



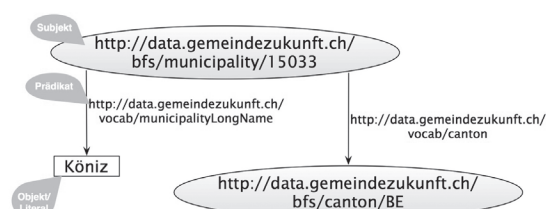
André Gerber



Sandro Kneubühl



Architektur des Testsystems auf `vmnabu.bfh.ch`



Beispiel eines RDF Graphen