

Graphdatenbanken für Historiker mit Perspektiven für die Historische Semantik

,
andreas.kuczera@geschichte.uni-giessen.de
Regesta Imperii Gießen/Mainz, Deutschland

Die zunehmende Mengen an Volltexten in den Geschichtswissenschaften und vor allem auch in der Mediävistik bietet neue Chancen für die Forschung, erfordern aber auch neue Methoden und Sichtweisen. Der Beitrag möchte die Verwendung von Graphdatenbanken für die Speicherung von Erschließungsinformationen vorstellen.

Momentan werden digitale Quellen und die mit ihnen verbundenen Erschließungsinformationen meist in XML oder in SQL-Datenbanken abgelegt. XML hat sich als Standard bewährt und findet in vielen Editionsprojekten als Datenformat Verwendung während Datenbanken auf Websites meist auf SQL-Datenbanken als Daten-Repositories zurückgreifen. XML-Dateien sind in der Regel bis zu einem gewissen Grade noch verständlich lesbar, bei SQL-Datenbanken ist die Lesbarkeit ohne Kenntnis der zu Grunde liegenden Datenstrukturen in der Regel nicht mehr gegeben. Dies liegt nicht zuletzt auch an den Architekturen der Datenbanken: um optimale Performance zu erhalten werden die Datenstrukturen normalisiert. Hier kommt es für die optimalen Nutzungsmöglichkeiten entscheidend auf die Gestaltung des Frontends der Datenbank an. Oft sind die User-Interfaces jedoch vor allem auf die Bedürfnisse jener Personen ausgerichtet, die die Datenbank selbst erstellt haben. Da diese Personen in der Regel die Datenstrukturen tief durchdrungen haben, kann es bei der Gestaltung des Frontends leicht zu einseitigen Ausrichtung auf Experten-Nutzer kommen. Solche Nutzer wissen bereits vor der Suchanfrage wie ihr Ergebnis aussieht. In den Fachwissenschaften wird eine solche Anfrage als CIN-Anfrage bezeichnet (concrete information need). Davon zu unterscheiden sind POIN-Anfragen (problem-oriented information need), bei denen der Nutzer ohne tiefere Kenntnisse des Datenmaterials und den zu Grunde liegenden Strukturen eine Anfrage startet (Vgl. hierzu Frants / Shapiro / Voiskunskii: 1997). Die Ausrichtung auf CIN-Anfragen zeigt sich auch in den größeren Quellenportalen zur Mediävistik (Vgl. Kuczera 2014). Hier ist die Verwendung von Graphdatenbanken ein alternativer Ansatz für die Speicherung von erschließendem Wissen. In SQL-Datenbanken sind die Informationen in Tabellen abgelegt, die untereinander verknüpft sind. Graphdatenbanken folgen hier einem völlig anderen Ansatz. In einem Graph gibt es Knoten und Kanten. Vergleicht man die Knoten mit einem Eintrag in

einer Tabelle einer SQL-Datenbank, wäre eine Kante eine Verknüpfung zwischen zwei Tabelleneinträgen. Im Unterschied zu SQL-Datenbanken können Knoten und Kanten jeweils Eigenschaften haben.

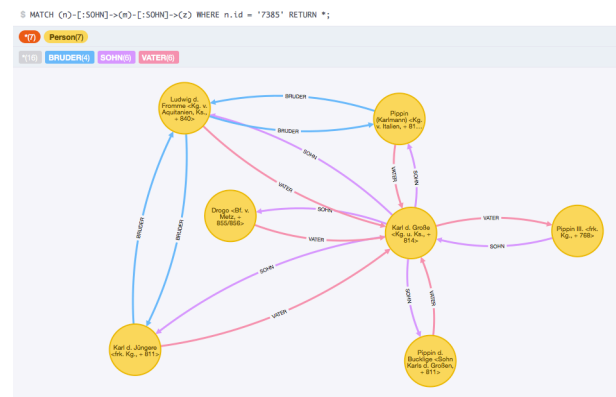


Abb. 1: Direkte Verwandschaftsverhältnisse Karls des Großen als Graph visualisiert

Daneben lassen sich die in Graphdatenbanken abgelegten Informationen sehr gut visualisieren. Gerade komplexere Datenbestände können hier sinnvoll für den Wissenschaftler erschlossen werden. Explorative Erschließungsmöglichkeiten erleichtern hierbei den Zugriff auf weitergehende Wissensdomänen des Repositoriums (Vgl. Kuczera 2015).

Das Datenmodell einer Graphdatenbank bildet quasi die semantische Repräsentation des in der Datenbank abgelegten Wissens. Ergänzt man die Eigenschaften der Knoten mit Identifikatoren wie den Angaben aus der GND oder legt man den Verknüpfungsstrukturen fachspezifische Ontologien zu Grunde können die Informationen der Graphdatenbank auch für automatisierte Abfragen über das Internet erschlossen werden.

In der Posterpräsentation werden in einem ersten Beispiel die Strukturen einer Graphdatenbank erläutert und anschließend mit der Graphenrepräsentation der Register der Regesten Kaiser Friedrichs III. und der genealogischen Datenbank Nomen-et-Gens Anwendungsbeispiele vorgestellt.

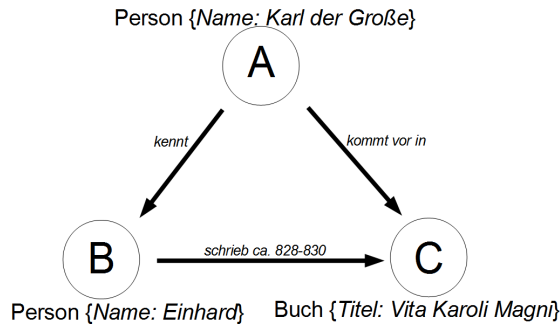


Abb. 2: Beispielgraph zu Karl dem Großen, Einhard und der Vita Karoli Magni

Bibliographie

Frants, Valery I. / Shapiro, Jacob / Voiskunskii, Vladimir G. (1997): *Automated information retrieval*. Theory and methods (= Library and information science). San Diego: Academic Press.

Kuczera, Andreas (2014): "Digitale Perspektiven mediävistischer Quellenrecherche", in: *mittelalter.hypotheses.org* <http://mittelalter.hypotheses.org/3492> [letzter Zugriff 28. September 2015].

Kuczera, Andreas (2015): "Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi", in: *mittelalter.hypotheses.org* <http://mittelalter.hypotheses.org/5995> [letzter Zugriff 28. September 2015].