

Leistungsfähige und einfache Suchen in lexikografischen Datennetzen Ein interaktiv-visueller Query Builder für Property-Graphen

Meyer, Peter

meyer@ids-mannheim.de

Institut für Deutsche Sprache, Deutschland

Einleitung: Property-Graphen für lexikografische Ressourcen

Klassische XML-basierte lexikografische Ressourcen können durch Graphenstrukturen mit zusätzlichen Vernetzungen und Informationen angereichert werden (M#chura 2016).¹ Dabei werden die Artikel eines Wörterbuchs zunächst durch eigenständige XML-Dokumente repräsentiert; bestimmte XML-Elemente in diesen Dokumenten – die in typischen Anwendungsfällen z.B. den im Artikel gebuchten Wörtern oder deren Bedeutungsdefinitionen entsprechen – können dann zusätzlich in einer Graphdatenbank für Property-Graphen (vgl. Robinson / Eifrem / Webber 2013) durch Knoten verschiedener Typen repräsentiert werden. Im Redaktionsprozess können auch z.B. weitere Knoten hinzugefügt werden, um zusätzliche Information abzubilden. Kanten zwischen solchen Knoten können nicht nur bereits vorhandene relationale Informationen aus den Quelldokumenten, sondern auch zusätzliche, insbesondere auch dokumentübergreifende, Relationen zum Ausdruck bringen. Der resultierende Graph fungiert dann als ausdrucksstarke zusätzliche Navigations- und Repräsentationsebene.

Ein Query Builder für die Graphensuche

Lexikograf/innen ebenso wie Endnutzer/innen einer solchen Ressource benötigen eine Zugriffsstruktur, die Suchen nach komplexen Konstellationen in solchen Graphen ermöglicht. Für Graphendatenbanken stehen zahlreiche Abfragesprachen zur Verfügung, deren Verwendung jedoch sehr voraussetzungsreich ist. Die Entwicklung von interaktiv-visuellen Systemen zur endnutzerfreundlichen Graphenabfrage ist aktuelles

Forschungsgebiet (vgl. z.B. Bhowmick / Choi / Li 2018; Pienta / Navathe / Tamersoy / Tong / Endert / Chau 2016).

Als eine auf die Bedürfnisse der digitalen Lexikografie zugeschnittene, sich insbesondere auch an interessierte Endnutzer sowie Lexikografen ohne IT-Vorkenntnisse richtende Lösung präsentiert das Poster einen visuellen Query Builder, der von den Komplexitäten der in vielen gängigen Property-Graphendatenbanksystemen implementierten Open Source-Abfragesprache Apache TinkerPop Gremlin (Rodriguez 2015; <http://tinkerpop.com>) abstrahiert.² Das Poster illustriert die Verwendung des Systems anhand einer Datenbank zu lexikalischen Entlehnungen aus dem Deutschen in andere Sprachen, die die mitunter verwickelten Entlehnungswege von Wörtern als Pfade in einem Graphen abbildet.

Abfragen werden im Browser durch das visuelle Zusammenstellen eines Baumes von Abfragekomponenten erzeugt, die Eigenschaften von Knoten beschreiben. Die so erstellten Abfragen sind zu jedem Zeitpunkt semantisch konsistent. Nach jeder Änderung an der Abfrage wird diese serverseitig in einer für Administratoren frei konfigurierbaren Weise in eine Gremlin-Graphtraversierungsanweisung umgesetzt und die Suchresultate in Echtzeit zurückgegeben. Angesichts der Mächtigkeit von Gremlin und der Möglichkeit, Traversierungen mit beliebigen Seiteneffekten zu verknüpfen, ist die direkte, manuelle Eingabe von Gremlin-Anweisungen nur in einer separaten, für Administratoren bestimmten Konsole möglich.

Im allgemeinen Fall werden n -Tupel von Knoten gesucht, die bestimmte Attribute aufweisen und zwischen denen nutzerdefinierte Pfade bestehen sollen. Entsprechend werden die Resultate tabellarisch als sortierbare n -Tupel präsentiert.

The screenshot shows a web-based Query Builder interface. On the left, there is a list of available properties (e.g., 'Sprache', 'Wortart', 'ist entlehnt als', 'Gruppe von Wortheigenschaften', 'Erstbelegjahr', 'hat Ableitung'). The main area contains a query tree where these properties are selected and combined with logical operators (AND/OR) and comparison operators (is, is not, is before, is after). The query is: 'Sprache ist Polnisch' AND 'Wortart ist Substantiv' AND 'ist entlehnt als ist Suchwort #1' AND 'Gruppe von Wortheigenschaften ist mindestens eine trifft zu' AND 'Erstbelegjahr ist nicht vor 1950' AND 'hat Ableitung ist Suchwort #2'. At the bottom, there are radio buttons to select the output format: 'nur Hauptsuchwort zeigen' (selected) or 'Tabelle mit allen Suchwörtern zeigen'. The results section shows two entries: 'szyńska POL' and 'шынська UKR'.

Abbildung 1. Beispiel für eine Query Builder-Suchanfrage in einem lexikografischen Netzwerk für Entlehnungsbeziehungen: Suche polnische Substantive, die als Lehnwort ohne Genuswechsel ins Ukrainische gewandert sind, wobei das Lehnwort oder eine Ableitung dazu nicht vor 1950 belegt ist.

Pienta, Robert / Navathe, Shamkant / Tamersoy, Acar / Tong, Hanghang / Endert, Alex / Chau, Duen Horng (2016): *"VISAGE: Interactive Visual Graph Querying"*, in: AVI: Proceedings of the Workshop on Advanced Visual Interfaces 272–279.

Robinson, Ian / Eifrem, Emil / Webber, Jim (2013): *Graph Databases*. Sebastopol, CA: O'Reilly & Associates.

Rodriguez, Marko A. (2015): *"The Gremlin Graph Traversal Machine and Language"*, in: **Cheney, James / Neumann, Thomas (eds.):** *Proceedings of the 15th Symposium on Database Programming Languages (DBPL 2015)*. New York: The Association for Computing Machinery 1-10.

Wood, Peter T. (2012): *"Query Languages for Graph Databases"*, in: SIGMOD Record 41(1): 50-60.