

Sechs Wege der FRBRisierung von Textverknüpfungen

Helling, Patrick

patrick.helling@uni-koeln.de
Universität zu Köln, Deutschland

Mathiak, Brigitte

mathiak@gmail.com
Universität zu Köln, Deutschland

Einleitung

Linked Open Data wird auch in den Geisteswissenschaften immer wichtiger (Barbera 2013: 91 – 105). Dabei geht es sehr oft um die Verknüpfung von Text mit weiterem Text, z.B. mit Kommentaren, Referenzen auf andere Werke oder anderen Entitäten wie Personen, Orte, etc. Mit Ontologien wie den Functional Requirements for Bibliographic Records object-oriented (FRBROO) (Bekiari et al. 2015) können zwar bibliographische Gegenstände im Rahmen einer digitalen Datenbank auf eine Art und Weise modelliert und bereitgestellt werden, dass diese allein durch ihre Organisation und objektübergreifende Strukturierung und Verknüpfung bibliographisch-informationellen Mehrwert im Rahmen der Recherche erzeugen können (Förster / Becker 2010: 15 - 25). Allerdings ist nicht immer klar, wie Verknüpfungen aus den Texten heraus mit anderen Texten oder aber zu komplett anderen Entitäten zu modellieren sind. Die Verknüpfung von verschiedensten Dokumenten und Entitäten ist allerdings eine der Hauptideen bei der Benutzung von Ontologien und Linked Open Data.

Einige Forschungsprojekte haben sich dieses Problems in Spezialfällen bereits angenommen. So wird bei HuCit (Romanello / Pasin 2011: 216 - 218) betrachtet, wie kanonische Zitation, z.B. der klassischen Literatur, modelliert werden kann. In (Bartalesi / Meghini 2016: 385 - 394) wird eine Ontologie speziell für die Texte von Dante Alighieri entwickelt. In (Mathiak / Boland 2015) wird der Spezialfall einer Verknüpfung zwischen Texten und Datensätzen, auf denen diese Texte beruhen, betrachtet.

Wir nehmen die zuvor genannten Ansätze als Ausgangspunkt, systematisieren diese und arbeiten Vor- und Nachteile heraus. Ziel ist es ein Modell für textbasierte Materialien zu konzipieren, mit dem Text(-stellen) mit anderen Textstellen oder Entitäten verknüpft und modelliert werden können.

Sechs Wege

1.) Die direkte Verbindung

In diesem einfachen Fall gibt es einfach eine Verbindung zwischen dem Dokument und der Entität, die mit diesem verknüpft werden soll. Die Property wird dazu idealerweise aus einer bereits etablierten Ontologie gewählt (siehe Abb. 1). Das klassische Beispiel hierfür ist eine Verschlagwortung oder die Zuordnung zu Autoren. Obwohl diese Art der Modellierung sehr einfach ist, hat sie doch starke Einschränkungen. Es bleibt unklar, welcher Teil des Dokuments für die Verbindung verantwortlich ist und es ist schwierig die Art der Verbindung über die Property hinaus zu beschreiben.



Abb. 1: Die direkte Verbindung.

2.) Die Verbindung über die Textstelle

Bei dieser Art der Modellierung wird zunächst die Textstelle innerhalb des Dokuments identifiziert und dann im nächsten Schritt mit der Entität verbunden, wodurch im Gegensatz zu Variante 1 eindeutig definiert wird, welcher Teil des Dokuments für die Verbindung verantwortlich ist (siehe Abb. 2).



Abb. 2: Die Verbindung über die Textstelle.

3.) Die Verknüpfung als eigene Entität

Dabei wird das Dokument bzw. die Textstelle zunächst mit einem Verknüpfungsobjekt verbunden und dann dieses mit der Zielentität (siehe Abb. 3), wobei im Vergleich zu Variante 1 die Erfassung relationsbeschreibender Informationen möglich wird. Diese Art der Modellierung ist vor allem nützlich, wenn es sehr viele Zusatzinformationen zu der Verknüpfung selbst gibt.

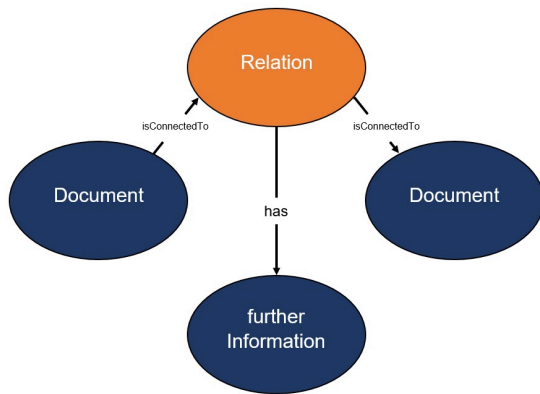


Abb. 3: Die Verknüpfung als eigene Entität.

4.) Die Verknüpfung mit einer Stellvertreterentität

Manchmal ist es nicht eindeutig möglich das Ziel der Verknüpfung als Entität zu identifizieren. In diesem Fall kann es sinnvoll sein, zunächst ein fiktives Ziel zu definieren und dann dessen Beziehungen zu bereits bekannten Entitäten zu etablieren (siehe Abb. 4). In gewissem Sinne ist dies invers zu Methode 2 zu sehen, bei der ja auch zunächst das Objekt der Textstelle neu erschaffen wird, indem es zu seinem Quelldokument in Beziehung gesetzt wird. In (Mathiak / Boland 2015) wird dies benutzt um fehlende Informationen zur Zielentität zu modellieren.

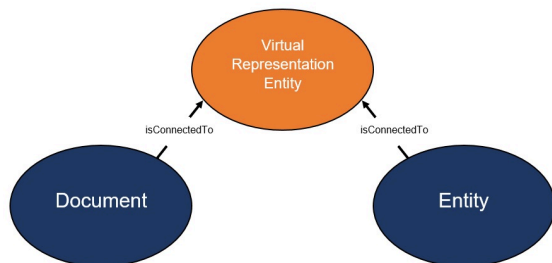


Abb. 4: Die Verknüpfung mit einer Stellvertreterentität.

5.) Das unabhängige Netzwerk

In vielen Fällen ist der Textbezug für die Ergebnisdarstellung nur von sekundärer Bedeutung, stattdessen werden die referenzierten Entitäten direkt miteinander in Verbindung gesetzt (siehe Abb. 5). Als Beispiel seien Beziehungsnetzwerke von Protagonisten in literarischen Werken genannt. Die Entitäten sind zwar auch mit dem Dokument verknüpft in dem sie auftauchen, aber die Kerninformation liegt in den Beziehungen, die diese untereinander haben.

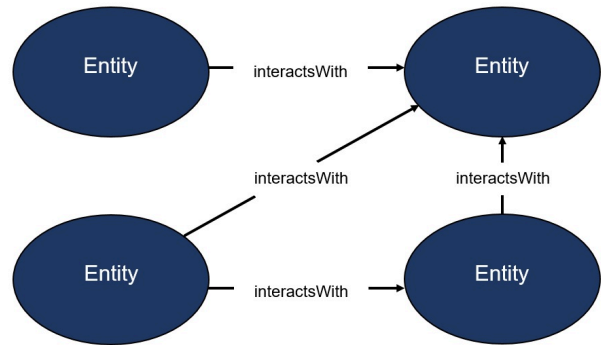


Abb. 5: Das unabhängige Netzwerk.

6.) Multiple Dokumente

Verschiedene Versionen desselben Dokuments werden in typischen Ontologien bisher wenig betrachtet, obwohl sie in den Digital Humanities eine häufige Modellierungsherausforderung sind.

Poster-Präsentation

Auf dem Poster präsentieren wir die grundlegende Ontologie-Struktur und stellen die verschiedenen modularen Modellierungsstrategien genauer vor. Zur Veranschaulichung werden wir eine Umsetzung unserer ontologischen Konzepte an einem Beispiel aus den Geisteswissenschaften demonstrieren.

Bibliographie

Barbera, Michele (2013): "Linked (open) data at web scale: research, social and engineering challenges in the digital humanities" in: *Global Interoperability and Linked Data in Libraries: Special issue. JLLS.it, Vol. 4, No. 1*: 91 – 105 <http://dx.doi.org/10.4403/jlls.it-6333>.

Bartalesi, Valentina / Meghini, Carlo (2016): „Using an Ontology for Representing Knowledge on Literary Texts: the Dante Alighieri Case Study“ in: *Semantic Web, Volume 8, Number 3, 6. Dezember 2016*: 385 – 394 <http://www.semantic-web-journal.net/content/using-ontology-representing-knowledge-literary-texts-dante-alighieri-case-study-0> [letzter Zugriff 22. September 2017].

Bekiari, Chryssoula / Doerr, Martin / Le Boeuf, Patrick / Riva, Pat (2015): „Definition of FRBRoo. A Conceptual Model for Bibliographic Information in Object-Oriented Formalism“ Den Haag: https://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo_v_2.4.pdf [letzter Zugriff: 25. September 2017].

Förster, Frank / Becker, Hans-Georg (2010): „Vernetztes Wissen – Ereignisse in der bibliographischen Dokumentation“ in: *Zeitschrift für Bibliothekswesen und Bibliographie*. 57. Jahrgang/Heft Nr. 1: 15 – 25 <http://dx.doi.org/10.3196/186429501057133>.

Mathiak, Brigitte / Boland, Katarina (2015): „Challenges in Matching Dataset Citation Strings to Datasets in Social Science“ in: *D-Lib Magazine*,

Volume 21, Number 1/2, Januar/Februar 2015, <https://doi.org/10.1045/january2015-mathiak> .

Romanello, Matteo / Pasin, Michele (2011): „An Ontological View of Canonical Citations “ *in: DH 2011 Book of Abstracts. Stanford: Stanford University Library* 216 - 218 <http://dh2011abstracts.stanford.edu/xtf/view?docId=tei/ab-143.xml> [letzter Zugriff 22. September 2017].