

**„Die Modellierung des Zweifels“.
Schlüsselideen und -konzepte zur
graphbasierten Modellierung von
Unsicherheiten**

Veranstalter: Akademie der Wissenschaften und der Literatur Mainz; Digitale Akademie Mainz; AG Graphentechnologien des Verbandes DHD

Datum, Ort: 19.01.2018–20.01.2018, Mainz

Bericht von: Simon Donig, Lehrstuhl für Digital Libraries und Information Systems, Universität Passau

Digitale Forschungsinstrumente haben in den letzten Jahrzehnten in allen historischen Disziplinen stetig an Bedeutung gewonnen. Datenbanken etwa sind längst von „einfachen“ Speicherlösungen zu komplexen Instrumenten der Wissensgenerierung geworden, deren Schemata selbst wiederum des methodenkritischen Blicks der Forschenden als Teil einer digitalen Hermeneutik bedürfen.

Zum zweiten Mal im Jahresabstand trafen sich Forschende aus dem Bereich der historisch arbeitenden digitalen Geistes- und Kulturwissenschaften, um unter dem Titel *Modellierung des Zweifels* die grundlegende Frage danach zu diskutieren, wie die Spezifika des Wissens in unseren Disziplinen digital mittels Graphen repräsentiert werden können und welche Rolle dabei Graphendatenbanken als Forschungsinstrument zukommen kann.

Seit einigen Jahren ergänzen Graphendatenbanken in Wissenschaft und Industrie relationale Datenbanken oder XML-basierte Forschungsdatenrepositorien oder lösen diese gar ab. Ihre enorme Flexibilität der Modellierung ist dafür nur ein Grund. Anders als die oftmals starren Schemata relationaler Repositorien verspricht der Graph eine Form der Wissensmodellierung, die den Anforderungen in den Geistes- und Kulturwissenschaften besser gerecht werden kann. Das betrifft insbesondere einen Kernbereich der historischen Disziplinen – den Umstand, dass viele Informationen, mit denen wir arbeiten, inhärent vage oder ungesichert sind und dass – verglichen mit klassischem *Big Data* – nur wenige Daten vorliegen, diese dafür aber in hochdimensionalen Datenräumen.

Die Konferenz hat genau diesen Umstand

zum Anlass genommen, die Frage danach aufzuwerfen, wie *Zweifel* mittels Graphen repräsentiert werden können. In seiner Einführung hob ANDREAS KUCZERA (Mainz) dann auch gerade auf die Frage ab, wie *Zweifel* gemessen und interoperabel gemacht werden könnten.

FRANZISKA DIEHR (Göttingen) schilderte dieses Problem für die Modellierung unsicherer Lesungen in einem digitalen Zeichenkatalog für die Schrift der klassischen Maya. Als polyvalentes Zeichensystem lassen die Maya-Glyphen mehrere Bedeutungen und zugleich sowohl Logogramm- als auch Silbenlesung zu. Der Herausforderung, mehrere Lesungsvorschläge zu fixieren, zu repräsentieren und zu testen, begegnet das Projekt, indem es nicht nur alle Lesungsvarianten repräsentiert, sondern auch verschiedene Vertrauensniveaus eingeführt hat, die auf formalen Kriterien beruhen. Damit wird es möglich, auf der Grundlage dieser Kriterien maschinell zu rasonieren. Die Vorteile der gewählten Lösung sah Diehr vor allem darin, dass die Kodierung des Korpus stabil bleibt, während alle Erkenntnisse über Lesung, Grammatik und ähnliches außerhalb des Korpus codiert werden. Der Graph unterstützt zugleich die Korpusprüfung und ermöglicht nicht zuletzt die Erkundung weitergehender Forschungsfragen, wie die nach Zeitabhängigkeit und räumlicher Verteilung bestimmter Zeichen.

Mit einer sehr ähnlichen Herausforderung sah sich MARTINA TROGNITZ (Wien) konfrontiert, die sich mit zweifelhaften Werten bei der Erschließung des *Corpus der Minoischen und Mykenischen Siegel* (CMS)¹ befasste. Forschungsfragen schließen jene nach der Regelmäßigkeit von Mustern im Arrangement der Zeichen, nach deren regionaler Verteilung, nach der Veränderung des Zeichenrepertoires über die Zeit und nach spezifischer Motivik ein. Unsicherheiten, berichtete Trognitz, lägen dabei auf mehreren Ebenen vor, da die Bedeutung der Zeichen nicht bzw. höchstens hypothetisch bekannt sei. Daneben bestünden Lesungsprobleme wie undefinierbare Zeichen. Eine besondere Herausforderung stellten weiter Zeichenkombinatio-

¹The CMS, URL <http://www.uni-heidelberg.de/fakultaeten/philosophie/zaw/cms/index_de.html>, Zugriff 03.02.2018

nen dar: Sagen etwa in einer Kombination aus drei Zeichen zwei bekannte die Sicherheit der Identifizierung eines unleserlichen Dritten voraus?

Ein übergreifendes Thema der Tagung war die Überführung bestehender Wissensrepositorien in eine Graphenform. In den textzentrierten Wissenschaften liegen die Ausgangsdaten häufig in XML/TEI vor. Der Musikwissenschaftler TORSTEN ROEDER (Würzburg) stellte Überlegungen zur Überführung von in XML/TEI transkribierten deutschsprachigen Musikkritiken am Beispiel von Giuseppe Verdis Requiem vor. Diese wurden anschließend im Graphen mit Normdaten angereichert. Neben Raumverteilungen, gegenseitiger Referenzierung von Texten untereinander (zum Beispiel Abgrenzung/Zustimmung zu anderen Kritiken) fragt das Projekt auch explizit nach Referenzen zu anderen musikalischen Werken. Letztere können in unterschiedlichen Graden von Vagheit bestehen („Verdi hat diesen Klangeffekt aus Beethovens 9. Symphonie übernommen“), sodass die letztlich identifizierten Referenzen in mehreren Varianten vorliegen können, die wiederum unterschiedlich plausibel sind. Weitere Unschärfen beinhalten Verweise auf ganze Werkgruppen sowie auf die Werke noch lebender Komponisten (Unabgeschlossenheit), die alle auch im Rahmen der Modellierung zu berücksichtigen seien.

Mit der Übertragung von Informationen aus XML/TEI, die bereits im Ausgangsmaterial als unsicher gekennzeichnet sind, beschäftigten sich DOMINIK KASPER (Mainz) und Andreas Kuczera. Nach einem Überblick über die Möglichkeiten für die Modellierung von Unsicherheit in XML/TEI (etwa: unleserlich, physisch lückenhaft) warfen sie vor allem die Frage nach *responsibility data* auf, also danach, wer mit welcher subjektiven Sicherheit eine Lesung beziehungsweise Ergänzung vorgenommen hat und wie dies im Graphen repräsentiert werden kann. Sie zeigten, wie sich anhand dieser Unsicherheitsbewertungen im Graphen schnell ein ‚editorischer Fingerabdruck‘ erstellen ließe. Kuczera betonte den produktiven Wert dieser Einschätzungen, da durch die Verbindung von subjektiven Sicherheitsangaben für eine Lesung mit Angaben über die Verantwortlichkeit die in-

tersubjektive Nachvollziehbarkeit des Editonsprozesses deutlich verbessert werde.

Nicht nur textorientierte Forschungen profitieren von Graphen als Repräsentationsform. Sie erlauben es auch, unterschiedliche Informationstypen miteinander zu verbinden und auf dieser Grundlage zu rasonieren (etwa georeferenzierte archäologische Befunde mit geographischen Angaben in althistorischen Texten).

MARTIN UNOLD und FLORIAN THIERY (beide Mainz) demonstrierten mit dem *Academic Meta Tool* (AMT)² ein Werkzeug zur Modellierung von Vagheit mittels *Reasoning* in der *Web Ontology Language* (OWL).

Unold betonte, dass Unsicherheit und Vagheit unterschiedliche Herausforderungen für das *Reasoning* darstellten. Er verwies zugleich darauf, dass lediglich Heuristiken zur Bestimmung von Vagheit möglich seien. Bei Vagheit stelle sich beispielsweise das Problem der Transitivität von Relationen wie „Stadt X liegt nördlich von Stadt Y, Stadt Y nördlich von Stadt Z. Folgt daraus, dass X ebenfalls nördlich von Z liegt?“. Im AMT werden die Knoten des Graphen grundsätzlich durch gewichtete Kanten verbunden, wobei die Gewichtung den Grad der Vagheit dieser Beziehung angibt. Mit Hilfe einer vordefinierten Ontologie kann auf der Grundlage der Kantengewichtungen dann maschinell rasoniert werden. Zugleich verfügt die Ontologie über 11 Axiome, mit denen automatisiert aus bestehenden neue Beziehungen zwischen den Knoten generiert werden können. Der Vortrag wurde durch ein Praxisbeispiel von FLORIAN THIERY und ALLARD MEES (Mainz) ergänzt, der mahnend mit auf den Weg gab, dass der Einsatz von graphenbasierten Repräsentationen kein modischer Selbstzweck sein dürfe.

Wie STEFAN MÜNNICH (Basel) in seinem Vortrag zu Quellenverlusten als methodologischem Unsicherheitsbereich für Editorik und Datenmodellierung zeigte, gibt es neben der Modellierung zwischen bekannten Informationen noch die Herausforderung der Modellierung von potentiell existenten Informationsobjekten. Am Beispiel von Anton Weberns *George-Liedern* zeigte er, wie die Exis-

² Academic Meta Tool. URL <<http://academic-meta-tool.xyz/>> (10.02. 2018).

tenz mehrerer früherer Fassungen einer Musikhandschrift aus Kontextannahmen (bisherige Arbeitspraxis des Komponisten) und einer weiteren Quellengattung (Briefe, persönliches Schrifttum) erschlossen werden kann. Bislang gebe es keine Arbeiten zu einem systematischen Modell von Quellenverlusten. Denkbar sei eine ganze Bandbreite solcher Verluste. Diese reichten von aus dem Bauchgefühl der Forschenden resultierende Annahmen über die Existenz von Quellen, über bekannte, aber physisch zerstörte oder in ihrem Verbleib ungeklärte Quellen bis hin zu fälschlichen Zuschreibungen. Damit das „quellenlose Vakuum“ (G. v. Dadelsen) überbrückt werden könne, bedürfe es eines Narrativs, also Hypothesen, die – um eine Untermodellierung zu vermeiden – expliziert werden müssten. Für eine Explizierung schlug er das *CRMinf Argumentation Model*, eine Erweiterung des CIDOC CRM³ vor.

Drei Ebenen von Mehrdeutigkeit und Unsicherheit unterschied ANDREAS WAGNER (Bern), indem er Unwissenheit (lückenhafte und unleserliche Vorlagen), vage und mehrdeutige Terme voneinander unterschied. Aus Sicht der Editionsphilologie seien insbesondere Emmendationen und Konjekturen noch nicht sinnvoll in Formalisierungen wie dem Argumentation Model des CIDOC CRM umgesetzt. In der anschließenden Diskussion wurde angeregt, die Feststellung von Unsicherheit nicht den subjektiven Einschätzungen der Forschenden zu überlassen, sondern beispielsweise Algorithmen zu verwenden, die die im Modell explizierten Unsicherheitskriterien als Input nähmen und so einen quantifizierbaren und replizierbaren Befund böten.

MICHAEL PIOTROWSKI, DAVIDE PICCA und MICHAEL MAIATSKY (alle Lausanne) vertieften die Frage nach der Abbildbarkeit und Modellierbarkeit von Unsicherheit in einer Weise, die (semi-) automatisiertem Rasonieren zugänglich ist. Der Vortrag griff die allgemeine Modelltheorie Herbert Stachowiaks auf⁴, deren Funktionsmerkmale der Abbildung, Verkürzung und pragmatischen Funktion er nutzte, um einen allgemeinen Rahmen der Modellbildung zu umschreiben. Begreife man die Narrative der Geisteswissenschaften als informelle Modelle, gelte es bei ihrer Ab-

bildung in einer Formalisierung Unsicherheit beziehungsweise argumentative Positionen, die nicht aufgelöst werden könnten, zu explizieren und damit greifbar zu machen. Ihre Mehrdeutigkeit müsse folglich angenommen (*accepted*) und produktiv modelliert werden. Dazu regte er die Nutzung von RDF und des *W3C Provenance Models*⁵ an. Diese parallel vorliegenden „*believe systems*“ mit verschiedener Autorschaft (Quellen) müssten dabei eben nicht notwendigerweise miteinander konsistent sein.

Eine beispielhafte Modellierung von Provenienz bot THOMAS EFER (Leipzig), der sich insbesondere mit Provenienzketten beschäftigte, die über nur einen oder wenige Knoten mit dem Graphen verknüpft sind. Er schlug vor, die Modellierung von Unsicherheit in den Begriffen Olaf Hartigs als Abstufungen zwischen *belief* und *disbelief* vorzunehmen.⁶

MARTIN RASPE (Rom) und GEORG SCHELBERT (Berlin) fügten der Tagung eine weitere Dimension hinzu, indem sie einerseits Unsicherheiten aus dem Bereich der Kunstgeschichte benannten, die noch stärker in den Bereich der Materialität und Modalität hinein reichten als die vorausgehenden Vorträge. Andererseits gingen viele ihrer Überlegungen vom an der Bibliotheca Hertziana angesiedelten Informationssystem Zuccaro⁷ aus. Für dieses, betonte Raspe, stehe der pragmatische Aspekt der Dokumentation und Repräsentation von Kunstwerken im Vordergrund, nicht die Nachbildung der Welt als Teil eines forscherschen Modells. Unsicherheiten seien

³CRMinf, URL <<http://new.cidoc-crm.org/crminf/home-4>> (03.02.2018).

⁴Herbert Stachowiak, Allgemeine Modelltheorie, Köln [u.a.] 1973.

⁵W3C Provenance Model, URL<<https://www.w3.org/TR/prov-overview/>> (02.02.2018); vgl. auch: Paolo Missier [u.a.], The W3C PROV family of specifications for modelling provenance metadata, in: Norman W. Paton [u.a.]: Proceedings of the 16th International Conference on Extending Database Technology, New York 2013, S. 773–776, URL <<https://openproceedings.org/2013/conf/edbt/MissierBC13.pdf>>(10.02.2018).

⁶Olaf Hartig, Querying trust in rdf data with *tsparql*, in: S. Bechhofer [u.a.] (Hrsg.), The Semantic Web: Research and Applications, Hamburg 2009, S. 5–20.

⁷Martin Raspe, Zuccaro - Ein modernes, konfigurierbares Informationssystem für die Geisteswissenschaften, 23.06.2014, URL <<http://zuccaro.biblherz.it/dokumentation/zuccaro>> (10.02. 2018).

in diesem Sinn vor allem Uneindeutigkeiten, schlechte immanente Beschreibbarkeit und die Einbettung der erfassten Artefakte in einem komplexen historischen Gefüge. Aus der Sicht der Kunstgeschichte gelte es daher, bislang wenig explizierte Konzepte wie das der Zuschreibung zu modellieren (etwa: eindeutig zugeordnet, zugeschrieben, unsichere Zuschreibung, sehr unwahrscheinlich). Raspe führte in diesem Zusammenhang aus, dass das gegenwärtige Zuccaro kaum *Reasoning* aus dem System selbst zulasse, und plädierte dafür, den Graphen explizit als augmentierten *Wissensgraphen* zu begreifen, der entsprechend *gemined* werden könne.

Die Tagung hat verdeutlicht, dass Graphen als Repräsentations- und Speicherform enorme Potentiale für die digitalen Geistes- und Kulturwissenschaften bieten. Sie sind jene Form der Wissensrepräsentation, die vielleicht am besten die Zeitgebundenheit und Hochdimensionalität von Informationen in den historischen Disziplinen akkommodieren kann. Zugleich bieten sie in ihrer interaktiven Traversierbarkeit ein eigenes Potential zur Wissensgenerierung, das unseren Disziplinen durch die Nähe zum etablierten Denk-konzept des Netzwerks nicht fremd ist.

Die rasche Entwicklung im Bereich des *Graph-Mining* in verschiedenen Nachbar-disziplinen wie der *Data Science*, darunter insbesondere die nächste technologische Welle der von künstlicher Intelligenz getriebenen Techniken der Wissensgenerierung, war auf der Konferenz allerdings höchstens in Ansätzen ein Thema. Die digitalen Geistes- und Kulturwissenschaften werden Wege, um sich diese Techniken produktiv zu eigen zu machen, und dabei auch neue Wege einer transdisziplinären Zusammenarbeit mit der Informatik erst noch finden müssen.

Konferenzübersicht:

Andreas Kuczera (Mainz): Begrüßung

Sitzung 1: Text

Sitzungsleitung: Thorsten Wübbena (Frankfurt am Main)

Torsten Roeder (Würzburg): Referenzielle Varianz in Texten über Musik

Franziska Diehr (Göttingen) et al.: Modellierung unsicherer Lesungen in einem digitalen

Zeichenkatalog für die Schrift der Klassischen Maya

Andreas Kuczera (Mainz) / Dominik Kasper (Mainz): Modellierung von Zweifel – Vorbild TEI im Graphen

Sitzung 2: Unsicherheit

Sitzungsleitung: Anna Neovesky (Mainz)

Martin Unold (Mainz) / Florian Thiery (Mainz): Academic Meta Tool – Ein Web-Tool zur Modellierung des Zweifels

Allard Mees (Mainz) und Florian Thiery (Mainz): Putting Samian pots together – modelling ceramic service family roots. Wie Graphen bei der Modellierung des Zweifels helfen können.

Stefan Münnich (Basel): Quellenverluste als methodologischer Unsicherheitsbereich für Editorik und Datenmodellierung am Beispiel von Anton Weberns George-Liedern op. 3 & 4

Sitzung 3: Theorie

Sitzungsleitung: Dominik Kasper (Mainz)

Manuel Villa (Washington D.C.): The Paradise Papers and neo4j

Michael Piotrowski (Lausanne) und Davide Picca (Lausanne): Accepting and Modeling Uncertainty

Andreas Wagner (Bern): 3 Levels of ambiguity and uncertainty

Sitzung 4: Technik

Sitzungsleitung: Thomas Kollatz (Essen)

Ian Neill (Glasgow): Certainty and uncertainty at the-codex.net

Katarina Adam (Berlin): Blockchain Technology und Graph Technology: Ist die Graphdatenbank die bessere Blockchain?

Lorenzo Speranzoni (Venedig) et al.: Unveiling Academic Collaboration Patterns Through Graph Databases

Sitzung 5: Erschließung

Sitzungsleitung: Andreas Kuczera (Mainz)

Martina Trognitz (Wien): Siegel mit sieben Siegeln – Zweifelhafte Werte zur Beschreibung von ägäischen Siegeln

Thomas Efer (Leipzig): Graphbasierte Model-

lierung von Faktenprovenienz als Grundlage für die Dokumentation von Zweifel und die Auflösung von Widersprüchen

Martin Raspe (Rom) und Georg Schelbert (Berlin): Genau, wahrscheinlich, eher nicht: Beziehungsprobleme in einem Wissensgraph

Tagungsbericht „*Die Modellierung des Zweifels*“. Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. 19.01.2018–20.01.2018, Mainz, in: H-Soz-Kult 08.05.2018.