

Exposé

**Formale, digitale Methoden in den
Geschichtswissenschaften. Am Beispiel digital
edierter Rechnungsbücher**

Christopher Pollin

19. April 2019, Graz

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Einleitung | 1 |
| 2 | Formale Methoden und Modelle in den Geschichtswissenschaften | 2 |
| 3 | Web of Data (aka Semantic Web) | 3 |
| 3.1 | Resource Description Framework | 3 |
| 3.2 | Ontologien | 6 |
| 3.2.1 | Vom Wissen, über das Semantische Netz zur Ontologie | 7 |
| 3.2.2 | Ontology Engineering | 10 |
| 3.2.3 | Reasoning | 10 |
| 4 | Historische Rechnungsbücher als Quelle | 10 |
| 4.1 | The George Washington Financial Papers | 10 |
| 4.2 | The Wheaton Accounts | 10 |
| 4.3 | Stagville Financial Papers | 11 |
| 5 | Assertive Digitale Edition von historischen Rechnungsbücher | 11 |
| 6 | Ein formales Modell für Rechnugsbücher: die Bookkeeping Ontology | 12 |
| 7 | Linked Open Data und Geschichte | 12 |
| 8 | Zusammenfassung | 12 |

1 Einleitung

Am Anfang der Arbeit steht neben einer allgemeinen theoretischen Diskussion zu klassischen Themen der Informationswissenschaft (Daten - Information - Wissen),¹ eine Auseinandersetzung auf konzeptioneller^{2,3} und technischer⁴ Ebene mit dem *Web of Data*, wobei ein Fokus auf Wissensmodellierung⁵, Ontologien⁶ und *Linked Open Data*^{7,8} liegt, sowie eine kritischer Auseinandersetzung mit dem *Web of Data*.⁹

Auf diesen Grundlagen und den Praxisbeispielen wird herausgearbeitet, dass es sich beim Web of Data nicht nur um einen Technologie Stack handelt, sondern um eine eigene Methode um wissenschaftliche Arbeit und Dokumentation¹⁰ zu befördern. Dazu werden zuerst formale Methoden in den Geisteswissenschaften diskutiert und anschließend um digitale Methoden erweitert.^{11,12} Das Ontology Engineering¹³ und das Reasoning¹⁴ spielen dabei eine hervorgehobene Rolle.

Der **Projektkontext** in dem ich angestellt bin, ist eine durch die *Andrew W. Mellon Foundation* geförderte und durch das *Wheaton College Massachusetts* koordinierte Kooperation des Zentrums für Informationsmodellierung mit Partnern aus den USA. Es verfolgt das Ziel semantisch angereicherte digitale Editionen von historischen Rechnungsbüchern einem breiten Fachpublikum zugänglich zu machen. Daten - aus unterschiedlichen Formaten - sollen auf einer gemeinsamen Plattform zusammengeführt werden und adäquate Formen des Retrievals, Discoverys und der Visualisierung eröffnen, um die Arbeit mit den Quellen zu erleichtern. Die Überführung nach RDF auf Basis der im Projektkontext entwickelten *Bookkeeping-Ontologie*, die Transferprozesse historischer Rechnungsbücher formalisiert, erlaubt die Interoperabilität, Verlinkung und Zusammenführung der Informationen im Sinne des *Web of Data* und *Linked*

¹FAVRE-BULLE, Bernard: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation. Springer, 2001.

²BERNERS-LEE, Tim/HENDLER, James/LASSILA, Ora: The semantic web. Scientific American, 284 2001 Nr. 5.

³CARDOSO, Jorge: The semantic web vision: Where are we? IEEE Intelligent systems, 22 2007 Nr. 5.

⁴BERNSTEIN, Abraham/HENDLER, James/NOY, Natalya: A New Look at the Semantic Web. Commun. ACM, 59 2016 Nr. 9 (URL: <http://doi.acm.org/10.1145/2890489>).

⁵KELLY, Richard: Practical knowledge engineering. Elsevier, 2016.

⁶STUCKENSCHMIDT, Heiner: Ontologien: Konzepte, Technologien und Anwendungen. Springer-Verlag, 2009.

⁷RIETVELD, Laurens et al.: Linked Data-as-a-Service: The Semantic Web Redeployed. In GANDON, Fabien et al. (Hrsg.): The Semantic Web. Latest Advances and New Domains. Cham: Springer International Publishing, 2015.

⁸BAUER, Florian/KALTENBÖCK, Martin: Linked open data: The essentials. Edition mono/monochrom 2011.

⁹SWARTZ, Aaron: Aaron Swartz's A Programmable Web: An Unfinished Work. Synthesis lectures on the semantic web: Theory and Technology, 3 2013 Nr. 2.

¹⁰LATOUR, Bruno: Zirkulierende Referenz. Bodentstichproben aus dem Urwald am Amazonas. 2016.

¹¹THALLER, Manfred: Digital Humanities als Wissenschaft. In Digital Humanities Springer, 2017.

¹²THALLER, Manfred: Ungefähre Exaktheit. Theoretische Grundlagen und praktische Möglichkeiten einer Formulierung historischer Quellen als Produkte ,unscharfer Systeme [1984]. Historical Social Research/Historische Sozialforschung. Supplement, 2017.

¹³HITZLER, Pascal et al.: Ontology Engineering with Ontology Design Patterns: Foundations and Applications. IOS Press, 2016.

¹⁴BURSZTYN, D. et al.: Reasoning on web data: Algorithms and performance. In 2015 IEEE 31st International Conference on Data Engineering. April 2015.

2 Formale Methoden und Modelle in den Geschichtswissenschaften

Am ersten August des Jahres 1808 hat ein gewisser James Haley 1/4 viertel Pfund Pulver, 1/4 Pfund Kugeln und einen 1/4 viertel Pfund Zucker zum Preis von 2 Schilling und 6 Pence im Laden der Stagville Plantage in North Carolina käuflich erworben. Solche Information über die Vergangenheit lassen sich in historischen Rechnungsbüchern finden. Dabei ist nicht der Einzeleintrag von großartiger Bedeutung für Fragestellungen in den Geschichtswissenschaften, sondern die Aggregation vieler Einzelinformationen, um eine Datengrundlage zu schaffen, auf der die Interpretation wie es denn gewesen sein könnte basieren kann.

In den 1980iger Jahren standen sich im methodischen Zugang zu solchen Quellen zwei Gruppen gegenüber, die sich als "Traditionalisten" und "Quantifizierer" festmachen lassen können. Im Gegensatz zu den "Traditionalisten", die einen hermeneutischen Zugang wählten, um historische Quellen zu verstehen, übernahmen die "Quantifizierer" formale Methoden, wie etwa statistische Verfahren, aus den Sozialwissenschaften, um sie auf Quellenkorpora anzuwenden und die daraus gewonnenen empirischen Fakten für die Interpretation zu nutzen.¹⁵ Ergebnis dieser Auseinandersetzung war es, dass bei der Anwendung formaler Methoden besonders auf die Nachvollziehbarkeit geachtet werden muss, damit nicht Dinge, die nicht empirisch Beweisbar sind, auch nicht so missverstanden werden können. Aus diesem Grund muss die Quelle ohne jegliche Vorannahmen zur Verfügung gestellt werden und einsehbar sein und die angewandte formale Modell bzw. die formale Methode in ihrer Gänze offen gelegt wird. Aber bei zweitem handelt es sich sowieso um den Kern der Geschichtswissenschaften: die Interpretationen der Vergangenheit, sollten für andere nachvollziehbar sein und wissenschaftlichen Kriterien standhalten. Latter is the core knowledge domain of historical research. It is advised to share the basic assumptions and definitions in a knowledge domain in a formal way.¹⁶

Nach dem abebben der formalen Methoden erleben sie zur Zeit eine neue Hochblüte in form von digitalen Methoden.

Manfred Thaller sprach bereits in den 1980 von einer knowledge domain.

Zentrale Concepte in diesem Zusammenhang sind das *Web of Data* bzw. *Semantic Web* und *Linked Open Data*. Diese Technologien ermöglichen es formale Modelle in den Geisteswissenschaften mensch und maschinenlesbar zusammen mit den Daten nachvollziehbar und nutzbar zu machen. Diese Technologien erleichtern den Austausch formaler Modelle. Über das Web können sie so leichter verteilt und nachgenutzt werden.

¹⁵JARAUSCH, Konrad Hugo/ARMINGER, Gerhard/THALLER, Manfred: Quantitative Methoden in der Geschichtswissenschaft. Wissenschaftliche Buchgesellschaft, 1985, S.XX-XX.

¹⁶THALLER, Manfred: Historical Information Science: Is there such a Thing? New Comments on an old Idea [1993]. Historical Social Research/Historische Sozialforschung. Supplement, 2017, S.XX-XX.

Um mein zweites Masterstudium (Geschichte) abzuschließen, soll sich die dafür notwendige Masterarbeit mit der Theorie und Anwendung formaler und digitaler Arbeitstechniken im Fachbereich Geschichte beschäftigen. Da ich in einem Projekt zur semantischen Anreicherung von digital edierten historischen Rechnungsbüchern in der technischen Umsetzung und Datenmodellierung angestellt bin bietet sich dieses Thema für eine Abschlussarbeit im Modul Historische Fachinformatik an.

Am Beginn dieses Exposés steht eine Spezifizierung der Thematik und Relevanz des Themas, sowie eine daraus abgeleitete Forschungsfrage. Nach einem Überblick über wichtige Literatur, Projekte und einer Skizze des Aufbaus der Arbeit, folgt eine Erörterung der Methoden und der Umsetzung. Am Ende werden erwartete Ergebnisse und Herausforderungen reflektiert.

3 Web of Data (aka Semantic Web)

Im Gegensatz zum klassischen Web, das als ein Web von Dokumenten betrachtet werden kann, versucht das *Web of Data* Daten aus unterschiedlichen Quellen zu integrieren und miteinander zu verknüpfen. Daten sollen so vorliegen, dass nicht nur Menschen diese in neuen Kontexten nutzen können, sondern auch Softwareagenten. Um diese Vision zu erreichen, die vom Erfinder des Web Tim BURNERS-LEE¹⁷ formuliert wurde, bedarf es der Umsetzung mehrerer aufeinander aufbauender technischer Grundlagen, die sich im *Semantic Web Stack* manifestieren. Auf dessen Basis, dargestellt in Abbildung ?? soll in diesem Kapitel die grundlegenden Technologien und Standards des *Web of Data* erörtert werden.

3.1 Resource Description Framework

Das *Resource Description Framework* (RDF) ist ein Datenmodell zur Darstellung und für den Austausch von Daten im Web. Daten werden in diesem Modell als Ressourcen definiert, wobei eine Ressource alles sein kann: ein Dokument, eine Person, ein physisches Objekt oder ein abstraktes Konzept. Über Ressourcen werden Statements der Form Subjekt-Prädikat-Objekt formuliert. Jedes Statement drückt eine Beziehung zwischen zwei Ressourcen aus. Das Subjekt und das Objekt stehen dabei für die beiden miteinander verbundenen Ressourcen; das Prädikat beschreibt die Art ihrer Beziehung. Diese Zusammensetzung von Subjekt, Prädikat und Objekt werden als Triples bezeichnet. Betrachtet man den Satz "Bob ist befreundet mit Alice", dann lässt sich folgendes Triple extrahieren: <Bob> als Subjekt, <ist befreundet mit> als Prädikat und <Alice> als Objekt. Ob Alice mit Bob befreundet ist geht aus diesem Statement noch nicht hervor, da jeder Relation in RDF nur eine Richtung definiert.¹⁸ In der graphischen Darstellung wird schnell klar, dass es sich beim RDF Datenmodell um einen gerichteten Graphen handelt,

¹⁷Vgl. BERNERS-LEE/HENDLER/LASSILA: Scientific American, Nr. 5, Bd. 284, 2001.

¹⁸Vgl. POWERS, Shelley: Practical RDF: solving problems with the resource description framework. O'Reilly Media, Inc., 2003, S.16-21.

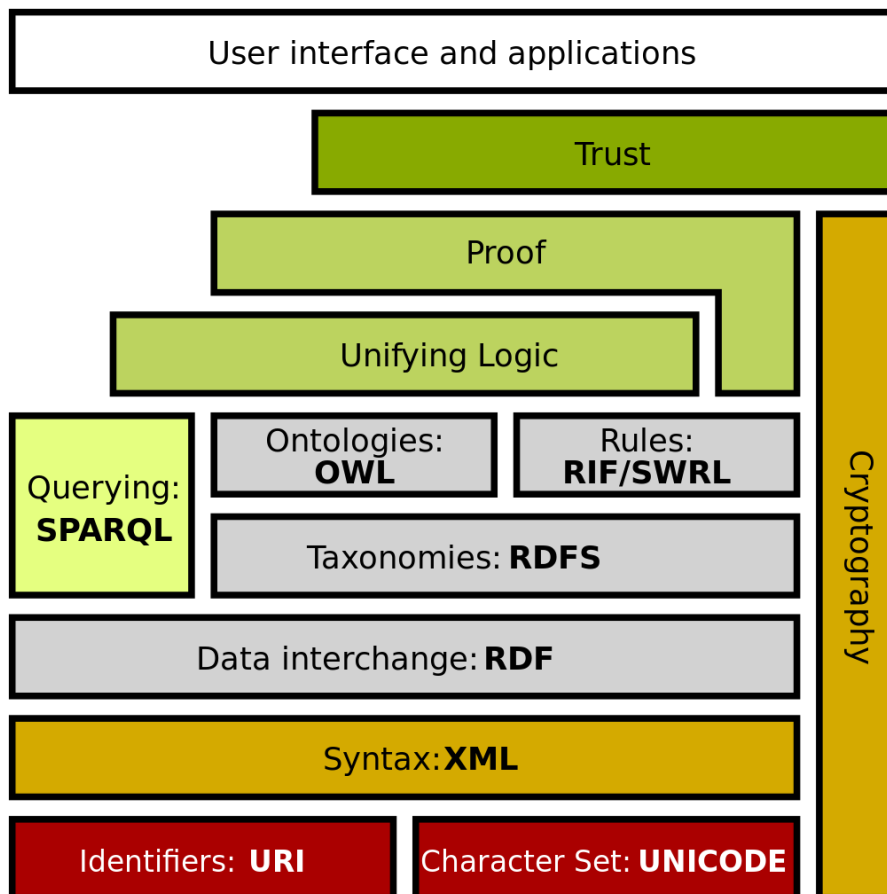


Abbildung 1: Visualisierung eines Graphen auf Basis eines RDF-Datensatz.

der aus Knoten (Subjekt und Objekt), sowie aus Kanten (Prädikat) besteht, wie Abbildung 2 zeigt.

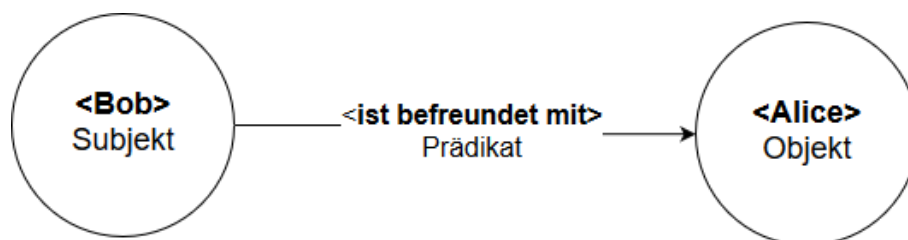


Abbildung 2: Semantic Web Stac

SCHREIBER und RAIMOND¹⁹ erklären RDF in einem ausführlichen Beispiel an Hand folgenden Aussagen:

Bob ist eine Person.

Bob ist befreundet mit Alice.

Bob ist geboren am 4. Juli 1990.

Bob interessiert sich für die Mona Lisa.

¹⁹Vgl. SCHREIBER, Guus/RAIMOND, Yves: RDF 1.1 Primer. W3C working group note 2014.

Die Mona Lisa wurde von Leonardo da Vinci entworfen.

Jede dieser Zeilen steht für ein Triple. *Bob* ist Subjekt in vier der oben genannten Tripeln, *Mona Lisa* tritt zweimal als Objekt und einmal als Subjekt auf. Dies ermöglicht es eine beliebige Menge an Triple zu einem komplexeren Graphen zusammenzusetzen und somit komplexere Sachverhalte beschreiben zu können. Abbildung 3 veranschaulicht das.

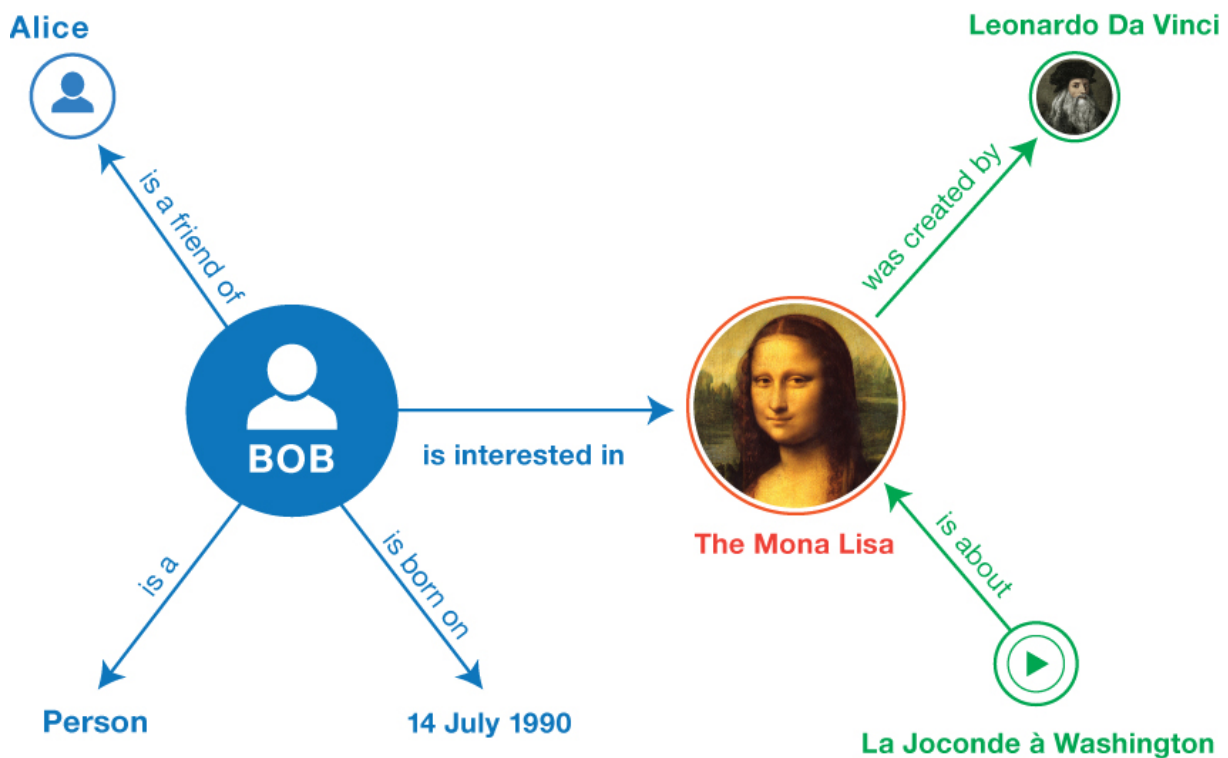


Abbildung 3: Visualisierung eines Graphen auf Basis eines RDF-Datensatz.

Uniform Resource Identifier (URI) können in allen drei Positionen eines Triple erscheinen. Somit ist jeder Ressource, sowie jeder Beziehung zwischen Ressourcen durch eine URI identifizierbar. URI's sind durch ein erweiterbares Schema definiert, damit Ressourcen im Internet eindeutig adressiert werden können. Um dabei die Einheitlichkeit zu gewährleisten, folgen sie einem vordefinierten Satz von Syntaxregeln, der 5 Komponenten beinhaltet:²⁰

$$URI = scheme:[//authority]path[?query][\#fragment]$$

- **scheme:** Definiert den Kontext und Typ. Bekannte Schemata sind beispielsweise die Webprotokolle *Hyper Text Transfer Protocol* (http) oder das *File Transfer Protocol* (ftp), sowie Notationskonzepte wie *Uniform Resource Name* (URN)urn oder *Digital Object Identifier* (doi).

²⁰Vgl. BERNERS-LEE, Tim/FIELDING, Roy/MASINTER, Larry: Uniform resource identifier (URI): Generic syntax. 2004 – Technischer Bericht (URL: <http://www.rfc-editor.org/info/rfc3986>).

- **authority**: Verwaltet Instanz in einem bestimmten vom Schema angegebenen Interpretationsraum, wie etwa das *Domain Name System*.
- **path**: Der Pfad enthält – oft hierarchisch organisierte – Angaben, die zusammen mit dem Abfrageteil eine Ressource identifizieren.
- **query**: Der Abfrageteil beinhaltet Daten zur Identifizierung von solchen Ressourcen, deren Ort durch die Pfadangabe allein nicht genau angegeben werden kann, wie beispielsweise ein Datensatz aus einer Datenbank, abgerufen werden
- **fragment**: Ist der optionale Fragmentbezeichner und referenziert eine Stelle innerhalb einer Ressource. Der Fragmentbezeichner bezieht sich immer nur auf den unmittelbar vorangehenden Teil des URI und wird von einem Hash (#) eingeleitet.

Weiters werden URI in *Uniform Resource Locator* (URL) und *Uniform Resource Name* (URN) unterteilt. Wo URN Namen von Ressourcen eindeutig identifizieren, wie etwa bei ISBN Nummern von Büchern, sind URL die gängigsten URI's, die den Ort einer Ressource adressieren und über einen Webbrowser auch aufrufen können.²¹ Für das Triple *<Bob> <interessiert sich für> <die Mona Lisa>* wird jeder Teilbestand eine URI und in der *Turtle* Serialisation von RDF ergibt es folgenden Code:

```

1 BASE <http://example.org/>
2 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
3 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
4 PREFIX schema: <http://schema.org/>
5 PREFIX dcterms: <http://purl.org/dc/terms/>
6 PREFIX wd: <http://www.wikidata.org/entity/>
7
8 <bob#me> a foaf:Person ;
9         foaf:knows <alice#me> ;
10        schema:birthDate "1990-07-04"^^xsd:date ;
11        foaf:topic_interest wd:Q12418 .
12
13 <wd:Q12418> dcterms:title "Mona Lisa" ;
14        dcterms:creator <http://dbpedia.org/resource/Leonardo_da_Vinci> .

```

3.2 Ontologien

Der Gegenstandsbereich der **Ontologie als Disziplin in der Philosophie** umfasst alles, das existiert. Das Erkenntnisziel, so MEIXNER, ist auf allgemeiner begrifflicher Ebene zu finden und

²¹Vgl. POWERS: Practical RDF: solving problems with the resource description framework, S.21-22.

beschäftigt sich mit der Einteilung des Seins und den Grundstrukturen der Wirklichkeit, sowie der Frage nach dem Wesen der Existenz. Die Ontologie verfolgt nicht das Ziel Erkenntnis über ein Objekt zu erhalten, es beispielsweise zu vermessen oder zu beschreiben, sondern stellt sich die Frage nach welchen allgemeinen Kriterien Objekte im Verhältnis zu ontologischen Begriffen wie Sein, Aktualität, Universalie, Exemplifikation, Sachverhalt oder Individuum stehen.²²

Der Begriff **Ontologie in der Informationswissenschaft bzw. Informatik** umfasst ein pragmatisches Konzept zum Austausch und zur Wiederverwendung von formalisierten und gemeinschaftlich verwendeten Wissensstrukturen durch ein gemeinsames Vokabular. Ziel dabei ist es Informationssysteme zu implementieren. Die Spezifikation eines solchen Vokabulars für eine bestimmte Domäne nennt man Ontologie.

Der Begriff wird in zwei Disziplinen mit jeweils unterschiedlichen Fokus verwendet. Dennoch sehe ich Gemeinsamkeiten. Beide setzen sich mit der Frage auseinander, wie die Welt sinnvoll strukturiert werden kann, damit wir uns besser darin zurecht finden können. In diesem Kapitel wird die informationswissenschaftlichen Dimension des Ontologie-Begriffs und seiner Nutzung in den digitalen Geisteswissenschaften diskutiert und der Frage nachgehen, ob Ontologien ein geeignetes Werkzeug zur Formalisierung von geschichtswissenschaftlichen Domänen darstellen. Dabei soll anfangs "Wissen" kurz aus informationswissenschaftlicher Sicht definiert werden und über das semantische Netz eine Brücke zur Ontologie geschlagen werden.

3.2.1 Vom Wissen, über das Semantische Netz zur Ontologie

Wissen ist eine systeminterne Repräsentation vorliegender Erfahrungen eines Menschen zu einem bestimmten Zeitpunkt, die einem zu überprüfenden Anspruch auf Gültigkeit ausgesetzt sein muss. Als solches prägt Wissen das Handeln und Denken eines Menschen auf den unterschiedlichsten Ebenen und dient zur Lösung von Problemen. Das jeweils aktuelle Wissen bildet einen kontextuellen Rahmen, in dem ankommende und bestehende Information interpretiert und zu neuen Erfahrungen verarbeitet werden.²³

Diese Definition von Wissen – eine stärker informationswissenschaftliche – hat seinen, neben vielen Definitionen in anderen Fachbereichen, legitimen Ursprung. Unterschiedlichen Disziplinen haben andere Fragestellungen und benötigen dafür ein anderes theoretisches Gerüst. Ein Wissensbegriff in der Philosophie, beispielsweise, sollte viel weiter gefasst sein, als ein Wissensbegriff in der Informationswissenschaft, dessen Aufgabe darin besteht als Hilfsmittel in der Entwicklung und Umsetzung von Informationssystemen zu fungieren.

Mittels Ontologie lässt sich "Wissen" als Netzwerk beschreiben. Ein Netzwerk ist ein gerichteter Graph, bestehend aus einer Menge von Knoten und einer Menge von Kanten, die die einzelnen Knoten miteinander verknüpfen. Damit lassen sich (fast) beliebige Entitäten und de-

²²MEIXNER, Uwe: Von der Wissenschaft der Ontologie. Logos (neue Folge), 1 1994.

²³FAVRE-BULLE: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation.

ren Verknüpfungen miteinander abbilden. Die Überlegungen zu einem **semantischen Netz**, als gedanklichen Vorgänger der Ontologie, stammen von QUILLIAN, der damit ein formales Erklärungsmodell für 'die menschliche Repräsentation von Wissen über Worte und ihre Bedeutung als Netzwerk von Begriffen und ihren Relationen'²⁴ beschreibt. Semantische Netze können einen Kompromiss zwischen menschenverständlicher Repräsentation einer Domäne und der formalen Verarbeitbarkeit durch eine Maschine darstellen.²⁵ Das ist dadurch gegeben, dass die Struktur des Graphen (=Netz), sich einfach in Rechnern als Matrizen abbilden lässt.

Die Ontologie ist eine Erweiterung des semantischen Netzes und nach GRUBER kann sie durch ein **4-Tupel** definiert werden. C ist eine Menge von **Klassen** (concepts, classes - Mengen von Entitäten aus der Realität), R eine Menge von **Relationen** (properties - Beziehungen zwischen Klassen), I eine Menge von **Instanzen** (individuals - einzelne Entität aus einer Menge) und A eine Menge von **Axiomen** (axioms - logische Regel).²⁶ C und R lassen sich dabei stets als Graph abbilden. Ein Beispiel zur Veranschaulichung:

Es existiert eine Klasse (C) "Katzen", die mit der Relation "ist ein"(R) mit einer Klasse "Säugetier" verbunden ist. Die Individuals (I) "Garfield" und "Tom" sind Instanzen der Klasse "Katzen" und erben alle Eigenschaften, die in der Klasse "Katzen" definiert wurden. Eine Regel kann definiert werden (A), sodass immer wenn eine Klasse eine "ist ein"-Verbindung zu einer Klasse wie "Säugetier" hat, es ausgeschlossen ist, dass es eine zweite "ist ein"-Verbindung gibt, die auf eine andere Klasse wie etwa "Vogel" referenziert.

Der Begriff der Ontologie terminologisch unscharf verwendet.²⁷ Die Unterschiede sind klein, aber dennoch entscheidend und sollen im Folgenden diskutiert werden. Eine der ersten Definitionen des Begriffs der Ontologie stammt von GRUBER:

*"An ontology is an explicit specification of a conceptualization"*²⁸

Eine "conceptualization" beschreibt den Prozess einer Vereinfachung, aber Fokussierung, eines bestimmten Aspekts der Realität. So kann eine Ontologie als Dokumentation eines wissenschaftlichen Prozesses agieren, in dem die Wirklichkeit abstrahiert und reduziert wird und gleichzeitig die Domäne bzw. Forschungsfrage hervorgehoben und amplifiziert wird.²⁹ Unter "explicit" versteht man, dass die Bedeutungen aller von der Ontologie erfassten Begriffe klar und eindeutig definiert sein müssen. Dies beinhaltet alle ihre Eigenschaften, Beschränkungen

²⁴STUCKENSCHMIDT: Ontologien: Konzepte, Technologien und Anwendungen.

²⁵REICHENBERGER, Klaus: Grundlagen semantischer Netze. In Kompendium semantische Netze Springer, 2010.

²⁶JOOST BREUKER, Pompeu CASANOVAS/KLEIN, MC/FRANCESCO, Enrico: The flood, the channels and the dykes: Managing legal information in a globalized and digital world. Law, Ontologies and the Semantic Web: Channelling the Legal Information Flood, 188 2009.

²⁷Vgl. GRUBER, Thomas: A translation approach to portable ontology specifications. Knowledge acquisition, 5 1993 Nr. 2, S.1.

²⁸HOEKSTRA, Rinke: Ontology Representation Design Patterns and Ontologies that Make Sense. In Proceedings of the 2009 conference on Ontology Representation: Design Patterns and Ontologies that Make Sense. Ios Press 2009, S.69.

²⁹Vgl. THALLER: Historical Social Research/Historische Sozialforschung. Supplement 2017.

und Beziehungen, innerhalb, als auch außerhalb der Domäne.³⁰ BORST erweitert GRUBERS Definition um '*formal specification of a shared conceptualization*'.³¹ "Formal"ergänzt dabei die Definition um die Notwendigkeit, dass Ontologien maschinenlesbar sein müssen. Erst diese Eigenschaft hebt sie von anderen Methoden zur Formalisierung von konzeptionellen Datenmodellen hervor. Der Zusatz "*shared*"reflektiert die Tatsache, dass eine Ontologie Wissen erfasst, das durch den Konsens einer Gruppe - z.B. durch einen wissenschaftlichen Diskurs - akzeptiert wird. Eine Ontologie darf nicht im Stillen von einer Person alleine entwickelt werden, sondern sollte in einem iterativen Prozess (Ontology Engineering) des Austausches und der Diskussion mit anderen entstehen. Ein solcher Prozess kann wie folgt ablaufen:

- Definition der Notwendigkeit und des Zieles einer Ontologie
- Strukturierung des Wissens und konzeptionelle Entwicklung
- Implementierung und Modellierung
- Evaluierung und Dokumentation
- Iteration dieser Punkte im Austausch mit anderen

Allgemeiner betrachtet definieren LINCKELS & MEINEL eine Ontologie als ein Datenmodell zur Darstellung eines Sets miteinander vernetzter Konzepte innerhalb einer (Fach-)Domäne.³² WELLER spricht von einer formalen und schematischen Darstellung einer Wissensdomäne auf Basis definierter Regeln und Vokabulars.³³

Zusammengefasst kann man sagen, dass sich mittels Ontologien komplexere Sachverhalte so darstellen lassen, dass Mensch und Maschinen in der Lage sind Strukturen, die durch eine Ontologie definierte und standardisierte sind, weiterverarbeiten zu können. Der Mehrwert kann vor allem in der Möglichkeit automatisierter Schlussfolgerungen, im Information Retrieval oder anderen formalen Methoden zur Verarbeitung von Daten liegen.

Der Ontology Editor Protégé erlaubt es, eine Ontologie und die darin enthaltenen Daten (Individuals) einem Reasoning - dem Abarbeiten aller Vorhanden Regeln in einer Ontologie auf Basis einer deskriptiven Logik - zu unterziehen. Für solche Zwecke gibt es natürlich auch API's und Bibliotheken in Programmiersprachen.³⁴ Das Reasoning gilt als ein essentieller Baustein im Design, der Entwicklung, der Wartung und in der praktischen Anwendung einer Ontologie. Das Ergebnis davon sind Inferenzen. Inferenzen sind neu hergeleitete Schlussfolgerungen

³⁰SURE, York/STUDER, Rudi: Methodology, tools & case studies for ontology based knowledge management. 2003.

³¹BORST, Willem Nico: Construction of engineering ontologies for knowledge sharing and reuse. 1997.

³²LINCKELS, Serge/MEINEL, Christoph: E-librarian service: user-friendly semantic search in digital libraries. Springer Science & Business Media, 2011.

³³WELLER, Katrin: Ontologien. In KUHLEN, Rainer/SEMAR, Wolfgang/STRAUCH, Dietmar (Hrsg.): Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 6. Auflage. Berlin and Bosten, 2013.

³⁴MUSEN, Mark A: The protégé project: a look back and a look forward. AI matters, 1 2015 Nr. 4.

auf Basis der formalen Regeln einer Ontologie.³⁵ Die Überprüfung strukturierter Daten mittels logischen Schlussfolgerungen kann dazu dienen, größere Datenmengen auf ihre Konsistenz und somit auch auf ihre Qualität hin zu prüfen, da logische Inkonsistenzen als Fehlermeldung angezeigt werden.

3.2.2 Ontology Engineering

3.2.3 Reasoning

4 Historische Rechnungsbücher als Quelle

Historische Rechnungsbücher liefern reichhaltige und strukturierte Datensätze, die oft längere Zeiträume abdecken und die als Aggregation vieler Einzelinformationen enthalten zu Beantwortung unterschiedlicher Forschungsfragen herangezogen werden können. Eine Transkription allein reicht nicht aus um die unterschiedlichen Dimensionen einer solchen Quelle abzudecken: die linguistische/textuelle, die quantifizierbare und die semantische Dimension. Für Forschungszwecke unterliegen historische Quellen einem Transformationsprozess hin zu (vernetzten) Informationsquellen, die in verschiedenen Forschungsszenarien genutzt werden können. Um dies zu veranschaulichen, werden drei Fallstudien von Projektpartnern und ihren jeweiligen Forschungsinteressen diskutiert, die weit über wirtschaftliche und administrative Aspekte hinausgehen.

4.1 The George Washington Financial Papers

The George Washington Financial Papers (1748-1799) ³ gives insight into the life of George Washington and other topics such as the material culture, social history, manufacturing and agriculture. The financial papers exist as digital edition, created and published via an open-source, Drupal 4 based editorial platform, and aim to make Washington's records freely accessible. The platform allows editing and publishing financial documents and gives the users the possibility to perform simple analytical functionalities. Samples of research questions that could be of interest to historians are: How much money did Washington spent annually and for which specific commodities? Which role slave trade plays in his business? How did the price of certain commodities fluctuate? What did the network of partners look like and who did business with him? How was the value of tobacco calculated through different currencies [St14]?

4.2 The Wheaton Accounts

The Wheaton Accounts (1828-1859) contain a daybook of a store selling commodities of daily life. The digital edition follows a TEI/XML approach. It extends the range of questions to

³⁵DENTLER, Kathrin et al.: Comparison of reasoners for large ontologies in the OWL 2 EL profile. Semantic Web, 2 2011 Nr. 2.

historical narratives and geographical information. It is interesting to follow an individual or a family as they appear in the daybook over time and reconstruct their social background for a historical narrative. The same applies to geographic information allowing to track geographic relationships of people or the origin of commodities [TB13].

4.3 Stagville Financial Papers

Das Digitalisierungs Projekt der Rechnungsbücher des Geschäftes der Plantage Stagville in North Carolina umfasst daybooks und ledgers aus den Jahren 1767-1892. Die Erschließung erfolgt durch einen open science und crowdsourcing Zugang über die Plattform *From the Page*, in dem die Quellen transkribiert und annotiert werden.

Forschungsfragen konzentrieren sich dabei auf die Anzahl und Verbindungen zwischen den einzelnen Käufern und Verkäufern, und welche Güter zusammen gekauft wurden. Weiters sind wirtschaftliche Abhängigkeiten, wer steht bei wem in der Schuld bzw. der soziale Status einer Person und ob sie dadurch anders kauft oder verkauft. Freie und Sklaven.³⁶

Ziel der Masterarbeit soll es nicht sein die Projektinhalte zu dokumentieren, sondern sich mit theoretischen und praktischen Fragestellungen zu formaler Modellen und formaler Methoden in den Geschichtswissenschaften auseinanderzusetzen, wiewohl Quellen, Workflows und Daten aus dem DEPCHA Projekt einfließen sollen.

5 Assertive Digitale Edition von historischen Rechnungsbüchern

Kapitel in dem digitale Edition, TEI und How to Bookkeep behandelt wird. Und die Besonderheiten editorische Arbeit mit Rechnungsbüchern.

TOMASEK und BAUMAN beschreiben ein Modell eines interpretativen Markups, um Beziehungen zwischen Individuen, Geld- Güter- und Dienstleistungstransfer, die Doppeleintrag-Buchhaltung umfassenm auszuzeichnen. Die Auszeichnung basiert auf den ausdrucksstarken Richtlinien der TEI.³⁷

³⁶BRUMFIELD, Ben/ANNA, Agbe-Davies: Encoding Account Books Relating to Slavery in the U.S. South. <https://medea.hypotheses.org/182>, 2015.

³⁷?, Vgl..

6 Ein formales Modell für Rechnungsbücher: die Bookkeeping Ontology

Im Zuge des Projektes **Digital Edition Publishing Cooperative for Historical Accounts (DEPCHA)**³⁸ wird ein gemeinsamer Publikations-Hub für historische Rechnungsbücher umgesetzt. Im Zentrum steht die Entwicklung und Nutzung einer Ontologie zur Formalisierung und Standardisierung von Buchungstransaktionen in historischen Rechnungsunterlagen, sowie ein *Linked Open Data* Zugang. Der Mehrwert entsteht durch Funktionalitäten des Retrievals, der Visualisierung und Analyse der eingespielten Datenbestände.

Für alle diese Projekte existieren hochstrukturierte RDF-Daten, die jeweils mit einer domänenspezifischen Ontologie beschrieben sind. Diese Ontologie wurde in einem iterativen *Ontology Engineering*-Prozess mit den FachkollegInnen, basierend auf den fachspezifischen Forschungsfragen, generiert.

Das *Data for History Consortium*³⁹ geht einen vergleichbaren Weg und versucht ein gemeinsames Set an Methoden im *Web of Data* zu entwickeln, um Daten in den Geschichtswissenschaften zu modellieren, verknüpfen und auszutauschen.

The common knowledge domain of these documents is formalized in a "bookkeeping" ontology, based on the REA model and compliant with the CIDOC CRM. As a conceptual data model, the ontology is developed in an iterative process. It formalizes the interpretation of transactions of money, commodities and services from one actor to another, and further properties that can be found in historical accounts. The RDF data extracted from the accounts becomes therefore a highly structured and self describing data set, being interoperable and reusable for researchers in diverse fields. The RDF representation can link to URI's of commodities, places, persons or other LOD vocabularies. Additionally the RDF representation contributes to the LOD. Thus, all formal methods applied in the DEPCHA project can be transferred to other data conforming to the proposed ontology and any kind of combined data set.

7 Linked Open Data und Geschichte

8 Zusammenfassung

³⁸gams.uni-graz.at/depcha

³⁹BERETTA, Francesco/BRUSEKER, George: The dataforhistory. org project: a proposal. In Workshop on the creation of an international Data for History consortium. 2017.

Literatur

- BAUER, Florian/KALTENBÖCK, Martin:** Linked open data: The essentials. Edition mono-/monochrom 2011
- BERETTA, Francesco/BRUSEKER, George:** The dataforhistory. org project: a proposal. In Workshop on the creation of an international Data for History consortium. 2017
- BERNERS-LEE, Tim/FIELDING, Roy/MASINTER, Larry:** Uniform resource identifier (URI): Generic syntax. 2004 – Technischer Bericht \langle URL: <http://www.rfc-editor.org/info/rfc3986> \rangle
- BERNERS-LEE, Tim/HENDLER, James/LASSILA, Ora:** The semantic web. Scientific American, 284 2001 Nr. 5, 34–43
- BERNSTEIN, Abraham/HENDLER, James/NOY, Natalya:** A New Look at the Semantic Web. Commun. ACM, 59 2016 Nr. 9, 35–37 \langle URL: <http://doi.acm.org/10.1145/2890489> \rangle
- BORST, Willem Nico:** Construction of engineering ontologies for knowledge sharing and reuse. 1997
- BRUMFIELD, Ben/ANNA, Agbe-Davies:** Encoding Account Books Relating to Slavery in the U.S. South. <https://medea.hypotheses.org/182>, 2015, Accessed: 2019-03-20
- BURSZTYN, D. et al.:** Reasoning on web data: Algorithms and performance. In 2015 IEEE 31st International Conference on Data Engineering. April 2015, 1541–1544
- CARDOSO, Jorge:** The semantic web vision: Where are we? IEEE Intelligent systems, 22 2007 Nr. 5, 84–88
- DENTLER, Kathrin et al.:** Comparison of reasoners for large ontologies in the OWL 2 EL profile. Semantic Web, 2 2011 Nr. 2, 71–87
- FAVRE-BULLE, Bernard:** Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation. Springer, 2001
- GRUBER, Thomas:** A translation approach to portable ontology specifications. Knowledge acquisition, 5 1993 Nr. 2, 199–220
- HITZLER, Pascal et al.:** Ontology Engineering with Ontology Design Patterns: Foundations and Applications. IOS Press, 2016
- HOEKSTRA, Rinke:** Ontology Representation Design Patterns and Ontologies that Make Sense. In Proceedings of the 2009 conference on Ontology Representation: Design Patterns and Ontologies that Make Sense. Ios Press 2009, 1–236

- JARAUSCH, Konrad Hugo/ARMINGER, Gerhard/THALLER, Manfred:** Quantitative Methoden in der Geschichtswissenschaft. Wissenschaftliche Buchgesellschaft, 1985
- JOOST BREUKER, Pompeu CASANOVAS/KLEIN, MC/FRANCESCONI, Enrico:** The flood, the channels and the dykes: Managing legal information in a globalized and digital world. Law, Ontologies and the Semantic Web: Channelling the Legal Information Flood, 188 2009, 3
- KELLY, Richard:** Practical knowledge engineering. Elsevier, 2016
- LATOURE, Bruno:** Zirkulierende Referenz. Bodentischproben aus dem Urwald am Amazonas. 2016
- LINCKELS, Serge/MEINEL, Christoph:** E-librarian service: user-friendly semantic search in digital libraries. Springer Science & Business Media, 2011
- MEIXNER, Uwe:** Von der Wissenschaft der Ontologie. Logos (neue Folge), 1 1994, 375–399
- MUSEN, Mark A:** The protégé project: a look back and a look forward. AI matters, 1 2015 Nr. 4, 4–12
- POWERS, Shelley:** Practical RDF: solving problems with the resource description framework. Ö'Reilly Media, Inc.", 2003
- REICHENBERGER, Klaus:** Grundlagen semantischer Netze. In Kompendium semantische Netze Springer, 2010, 3–19
- RIETVELD, Laurens et al.:** Linked Data-as-a-Service: The Semantic Web Redeployed. In **Gandon, Fabien et al. (Hrsg.):** The Semantic Web. Latest Advances and New Domains. Cham: Springer International Publishing, 2015, 471–487
- SCHREIBER, Guus/RAIMOND, Yves:** RDF 1.1 Primer. W3C working group note 2014
- STUCKENSCHMIDT, Heiner:** Ontologien: Konzepte, Technologien und Anwendungen. Springer-Verlag, 2009
- SURE, York/STUDER, Rudi:** Methodology, tools & case studies for ontology based knowledge management. 2003
- SWARTZ, Aaron:** Aaron Swartz's A Programmable Web: An Unfinished Work. Synthesis lectures on the semantic web: Theory and Technology, 3 2013 Nr. 2, 1–64
- THALLER, Manfred:** Digital Humanities als Wissenschaft. In Digital Humanities Springer, 2017, 13–18

- THALLER, Manfred:** Historical Information Science: Is there such a Thing? New Comments on an old Idea [1993]. Historical Social Research/Historische Sozialforschung. Supplement, 2017, 260–286
- THALLER, Manfred:** Ungefähre Exaktheit. Theoretische Grundlagen und praktische Möglichkeiten einer Formulierung historischer Quellen als Produkte ,unscharfer ‘Systeme [1984]. Historical Social Research/Historische Sozialforschung. Supplement, 2017, 138–159
- WELLER, Katrin:** Ontologien. In **Kuhlen, Rainer/Semar, Wolfgang/Strauch, Dietmar (Hrsg.):** Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. 6. Auflage. Berlin and Bosten, 2013, S.207–218

Abbildungsverzeichnis

| | | |
|---|---|---|
| 1 | Visualisierung eines Graphen auf Basis eines RDF-Datensatz, https://www.w3.org/TR/rdf11-primer/ , 10.04.2019. | 4 |
| 2 | Semantic Web Stac | 4 |
| 3 | Visualisierung eines Graphen auf Basis eines RDF-Datensatz, https://www.w3.org/TR/rdf11-primer/ , 10.04.2019. | 5 |