

Prä- versus postkoordinierende Ontologien

Johannes Busse¹

Abstract: Unsere Leitfrage lautet: *Was stelle ich als Dozent mir vor, wenn ich den Studierenden zeigen will, was eine Ontologie „ist“?* Folgende Position wird argumentativ entwickelt: Besonders interessant ist es, sich nicht einen Baum (oder DAG etc.) von Klassen vorzustellen, sondern *eine Sammlung von Unterschieden, die einen Unterschied machen*. Denn es macht einen Unterschied, ob man nach Gemeinsamkeiten sucht oder ob man mögliche Unterschiede zu systematisieren versucht. Um in der Lehre den Aufbau und die Verwendung von Ontologien praktisch zu unterrichten, sollte eine Kunst des probleminduzierten Unterscheidens entwickelt werden. Als Werkzeug und Notation bietet sich GenDifS an, da sich hier Unterschiede expliziter unterscheiden lassen als in anderen Kombinationen von Ontologie-Sprache und -Visualisierung.

Keywords: Ontologie, Linked Open Data, Selbstexploration, GenDifS

1 Einführung

Im zweiten Semester in unserem neuen Bachelorstudiengang Verwaltungsinformatik unterrichte ich das Modul *Linked Open Government Data*, in dem insbesondere auch Kompetenz bzgl. des nationalen Metadatenportals govdata.de vermittelt werden soll. GovData verlangt von den Datenbereitstellern, dass sie ihre Daten „im RDF-Format und konform zum geltenden Metadatenstandard [DCAT-AP.de](https://www.govdata.de/dcat-ap)“ bereitstellen ([govdata.de > sparql-assistent](https://www.govdata.de/dcat-ap)). Dreh- und Angelpunkt dieses Technologiestapels sind Semantic Web Ontologien. Für die Verschlagwortung von Datensätzen ist die Thesaurus-Ontologie SKOS Standard, die auch von GovData, dem EU-Thesaurus EUROVOC sowie dem schon in GovData erwähnten Metadaten-Standard [\[DCAT22\]](https://www.govdata.de/dcat-ap) referenziert wird. Nach einem Hype vor gut 20 Jahren und einem Tal der Enttäuschung vor gut 10 Jahren hat sich das Semantic Web inzwischen offensichtlich als Standard für die Datenmodellierung durchgesetzt.

Wenn Studierende einer Bindestrich- (hier: Verwaltungs-)Informatik die Zusammenarbeit zwischen technischen Dienstleistern und ihrem Auftraggeber verantworten wollen, müssen sie die grundlegenden Prinzipien von RDF(S), Metadatenmodelle, Semantic Web etc. verstehen. Auch für Studierende nicht primär technischer Fächer wird offensichtlich eine Einführung in das Semantic Web und Ontologien benötigt. Anders als bei Informatik- oder KI-Studiengängen können wir bei diesen Adressaten nicht auf formal-logisches Vorwissen aufbauen, sondern müssen auf konkrete, einschlägige Ontologien zurückgreifen, die in der Praxis eine Rolle spielen.

¹ HAW Landshut, Fakultät Informatik, HAW Landshut, Am Lurzenhof 1, 84036 Landshut
busse@haw-landshut.de, <https://orcid.org/0000-0002-7051-6198>

2 Leitende Fragestellung

Der vorliegende Aufsatz diskutiert nicht die Didaktik einer solchen Einführung, sondern dient der Vergewisserung, was das für ein Ding ist, in das eingeführt werden soll. Insbesondere soll eine systematische Reflexion unter Kollegen angeregt werden zur Frage: *Was stellen wir als Ontologen; und was stelle ich als Dozent mir eigentlich vor, was eine Ontologie „ist“?* Obwohl ich mich im Bereich Semantic Web Ontologien als Experte bezeichnen würde, verfüge ich doch nicht über ein umfassendes Modell verschiedener Sichtweisen auf den Begriff der Ontologie. Umso mehr muss ich, wenn ich in meiner Lehre in Ontologien einführen will, zumindest selbst über ein einigermaßen konsistentes Verständnis verfügen, wohin der Unterricht gehen soll.

3 Methodologie

Im vorangehenden Text wird ein Kernthema dieses Aufsatzes aus der Ich-Perspektive vorgestellt: *Umso mehr muss ich, wenn ich in meiner Lehre in Ontologien einführen will, zumindest selbst über ein einigermaßen konsistentes Verständnis verfügen, wohin der Unterricht gehen soll.*

Die Ich-Perspektive mag ungewöhnlich erscheinen, sie irritiert und gilt unter nicht wenigen Kollegen als unwissenschaftlich. Sie ist aber in diesem Aufsatz methodisch gewollt und bedarf einer besonderen Thematisierung. Weil philosophische Methodendiskussion ein wichtiger Teil einer jeden Wissenschaft ist, wollen wir diese Perspektive gemeinsam reflektieren.

Im einfachsten Fall handelt es sich bei der Ich-Form lediglich um eine Syntax, die Passivsätze vermeiden will: Was versteht man / was verstehen wir / was verstehe ich unter X? Denkbar ist die Ich-Form natürlich auch als Ausdruck von Narzissmus, aber auch als Gegenteil von Unsicherheit oder Bescheidenheit. Eine mit *IMHO* (in my humble opinion) getaggte Nachricht gibt eben nicht vor, dass es sich hier um sicheres oder maßgebliches Wissen handele. Der Ich-Bezug zeigt vielmehr an, dass hier nur eine von vielen möglichen Meinungen vorliegt. In all diesen Fällen würde sich das „ich“ syntaktisch durch ein „wir“ oder „man“ oder durch Passivkonstruktionen ersetzen lassen.

In diesem Aufsatz trifft all das nicht zu. Tatsächlich will der Autor eigene Vorstellungen reflektieren. Nicht primär aus narzisstischen Gründen und auch nicht primär aus einer selbstkritischen „IMHO“-Einstellung gegenüber intersubjektiven Erkenntnismöglichkeiten. Sondern weil auch in der Design Science Forschung (DS Research, DSR) die Subjektivität der Erkenntnis – auch wenn sie mit Objektivität oder wissenschaftlichem Rigor in einem Spannungsfeld zu stehen scheint – eine bedeutende Rolle spielt. Wir wollen diese Spannung in diesem Methodenteil genauer beleuchten.

Laut Hevner gehören zu den Rahmenbedingungen einer Problemlösung im Rahmen der Design Science Forschung (DS Research, DSR) die “goals, tasks, problems, and opportunities that define business needs”, wie sie durch die “roles, capabilities, and characteristics of people within the organization” ([HMP+04], p.81) wahrgenommen und beschrieben werden. Schon der Problemraum in der DSR ist also ganz wesentlich durch subjektive Elemente bestimmt. Aber auch der DS-Researcher selbst ist kein außenstehender, sondern ein teilnehmender Beobachter, dessen eigenes Verstehen als konstitutiver Teil zur Problemdefinition beiträgt. Es sind auch seine persönlichen, subjektiven Eigenschaften, die wesentlich zur Problemlösung beitragen:

Design-science research in IS addresses what are considered to be wicked problems [...] those problems [are] characterized by [...] a critical dependence upon human cognitive abilities (e.g., creativity) [...] and] human social abilities (e.g., teamwork) to produce effective solutions. ([HMP+04], p.81)

Den subjektiven, kreativen, sozial eingebundenen, handelnden und verantwortungsfähigen Menschen im wissenschaftlichen Prozess unsichtbar zu machen, ist mit Bezug auf Hevner nicht zielführend. Es ist gut und richtig, in der Wissenschaft Objektivität einzufordern. Man würde aber das Kind mit dem Bade ausschütten, wenn man methodisch systematisch ausblenden würde, was Subjektivität, Erleben, Erfahrung, Kreativität, Gefühle und moralische Bewertungen zum Erkenntnisprozess beitragen. Besser ist es, auch solche subjektiven Elemente methodisch adäquat zu begleiten.

Auch wenn uns unser Bauchgefühl sagt, dass selbst ein unbestimmtes Bauchgefühl gelegentlich einen berechtigten Platz in der Forschung hat, fühlen sich Wissenschaftler in Bezug auf Bauchgefühl zu Recht unwohl. Die spannende Frage lautet also, wie sich der Geltungsanspruch menschlicher Lösungsbeiträge mit dem in der Design Science Forschung so energisch eingeforderten Rigor reflektieren lässt.

Eine anerkannte Methode ist die des Diskurses. In einem Dagstuhl-Forschungstreffen ([Bu23b]) hatten sich knapp 20 Menschen aus Praxis und Wissenschaft in einem offenen, transdisziplinären Workshop-Format zusammengefunden, das insbesondere dazu diente, nicht Antworten und Ergebnisse zu präsentieren, sondern in fragender Haltung Ansatzpunkte des Zweifels an Sichergeglaubtem zu finden. Das Ziel bestand darin, *neue* Fragen zu stellen. Einer der anspruchsvollsten Aufgaben innerhalb unserer Diskurse bestand darin, die eigenen wissenschaftlichen Voraussetzungen und Vorurteile sichtbar zu machen und vorausgehend diese überhaupt einmal zu erkennen. Angestoßen werden solche Prozesse z. B. durch Fremdheitserfahrungen, Nichtverstehen oder kognitive Dissonanzen, wie sie vielleicht nur bei Transdisziplinarität vorkommen.

Eine wichtige Voraussetzung in Diskursen, die geeignet sind, die Grundlagen des eigenen Faches zu überprüfen und kritische Anfragen nicht schon frühzeitig zurückzuweisen, ist Ambiguitätstoleranz. Eine wichtige Methode, die eigene Meinung einer solchen Überprüfung zu unterziehen, ist die Selbstexploration, wie sie z. B. auch in der klientenzentrierten Psychotherapie nach Carl R. Rogers als zentrale Dialogform eingesetzt wird, nämlich ein

„Aussprechen (verbalisieren) des eigenen inneren Erlebens und der gegenwärtigen Erfahrung sowie der damit verbundenen Gefühle und Bewertungen einer Person“ [Wikipedia > Selbstexploration](#).

Verbalisieren vom innerem Erleben und Gefühlen durch Selbstexploration: Als wissenschaftliches Element in der Wirtschaftsinformatik scheint dies tatsächlich weit hergeholt. Und doch gibt uns gerade auch Hevner gute methodologische Gründe an die Hand, auch solche Gedankenexperimente nicht dogmatisch auszuschließen, sondern als Möglichkeit der Weiterentwicklung der DSR zu erkennen:

[...] As a result [...] a theory of design in information systems, of necessity, is in a constant state of scientific revolution (Kuhn 1996). ([HMP+04], p.81)

Wer als Strukturwissenschaftler im Bereich Ontologien unterwegs ist, wird feststellen, dass man es nicht nur mit belanglosen Glasperlenspielen zu tun hat, sondern dass ein enger Zusammenhang besteht zwischen den Modellen der formalen (hier Semantic Web-)Ontologien und der eigenen Interpretation und Sicht auf Welt. Ontologien sind kognitiv wirksam. Als oberste Meta-Modelle sind sie das Mittel der Wahl z. B. zur semantischen Datenintegration. Gleichzeitig sind oberste Meta-Modelle nun einmal dadurch definiert, dass sich nur noch bedingt höhere Meta-Meta-Modelle finden und formal erfassen lassen, die man bei einem Konflikt von Sichtweisen auf Welt als gemeinsame Sprache heranziehen könnte. Umso wichtiger ist es für Semantic Web Ontologen, immer wieder neu zu reflektieren, welche philosophisch-ontologischen Implikationen unsere Semantic Web Ontologien haben, und sich dazu grundlegend erst einmal selbst klar zu werden, was man selbst sich unter einer Ontologie vorstellt.

Damit ist auch der wissenschaftliche Ertrag dieses Methodenteils angedeutet: Die Beschäftigung mit Ontologien erfordert es, auch Selbstexploration in den Methodenkanon der DSR aufzunehmen.

4 Stufen ontologischer Präzision

Wer über Ontologien schreibt, zitiert typischerweise zuallererst Tom Gruber: “An ontology is an explicit specification of a conceptualization” ([Gr95]). Das ist sozusagen die Arbeitsdefinition, die die Szene zusammenhält. Pragmatiker sind mit dieser Definition zufrieden und bereit, jede in der Praxis hilfreiche Wissensrepräsentation (z. B. Glossar, Terminologie, Thesaurus, Taxonomie, Wortnetz) zumindest als „leichtgewichtige“ Ontologie zu bezeichnen.

Dogmatiker – zu denen sich auch der Autor dieses Textes zählt – interessieren nur „schwergewichtige“, voll formalisierte Ontologien, aus denen eine anspruchsvolle formale Logik nicht hinwegzudenken ist, ebenfalls mit Berufung auf Gruber:

In such an ontology, definitions associate the names of entities in the universe of discourse [...] with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms. Formally, an ontology is the statement of a logical theory.[fn 1:] To specify a conceptualization one needs to state axioms that *do* constrain the possible interpretations for the defined terms. ([Gr95], p.2, Hervorhebung Gruber).

[Bi21] differenziert kenntnisreich und detailliert verschiedene Typen von Ontologien.

5 **Ontologie-Visualisierung: Baum, DAG, Semantisches Netz?**

Um eine Ontologie nutzbar zu machen, müssen ihre Strukturen geeignet visualisiert und die Ontologie mit geeigneten Werkzeugen erstellt, gepflegt, auf Konsistenz überprüft, in abgeleitete Wissensrepräsentationen exportiert und anderen Applikationen zur Verfügung gestellt werden. In der Semantic Web Community wird typischerweise das Experten-Tool Protégé ([Mu15]) genutzt. Protégé hat verschiedene Visualisierungs-Plugins (Google Image Search > “protege visualization plugin”), die in den meisten Fällen allerdings nur Views liefern und somit nicht editierbar sind. In der GUI des Editors ist die Baum-Ansicht im linken Fenster dominant, mit der Klassen ausgewählt werden, um sie dann in den anderen Fenstern im Detail editieren zu können. Bei einer Mehrfach-Vererbung kommt dieselbe Klasse an verschiedenen Stellen im Baum vor – eine suboptimale Lösung, die nicht skaliert und schnell unübersichtlich wird.

Ontologien als Baum zu visualisieren, funktioniert in den oberen Ebenen einer Ontologie meist recht gut. Da aber in praxisrelevanten Ontologien in tieferen Schichten rasch auch Mehrfach-Vererbungen auftreten, skaliert die Baum-Metapher nicht, sondern wir müssen mit gerichteten, azyklischen Graphen (directed acyclic graph, DAG) als Visualisierung arbeiten.

Bisweilen werden Ontologien auch als ein Semantisches Netz visualisiert, so z. B. auch im (eher irreführenden als hilfreichen) Artikel [https://de.wikipedia.org/wiki/Ontologie_\(Informatik\)](https://de.wikipedia.org/wiki/Ontologie_(Informatik)). Stuckenschmidt zeigt, wie schwierig die Überführung dieses Beispiels in eine formale Logik ist, die tatsächlich praktisch verwendbar ist ([St11], S.30ff, 40-51). Dennoch sind semantische Netze in der Praxis als informelle Wissensrepräsentationen extrem relevant und unter bestimmten Bedingungen tatsächlich auch als Visualisierung einer Ontologie geeignet.

Aus der Sicht eines Dogmatikers ist eine Ontologie einerseits ein formales System, eine formal-logisch axiomatisierte Terminologie. Andererseits gibt es in der Praxis viele Wissensrepräsentationen, die eben *nicht* voll formalisiert werden dürfen, um praktisch relevant zu werden: Eine Terminologie, ein Wortnetz, ein Thesaurus, ein Semantisches Netz,

ein Knowledge Graph leben oft von ihrer semantischen Unbestimmtheit und Interpretationsoffenheit. Der praktische Zugang zur Wissensrepräsentation geschieht über Repräsentationen der Wissensrepräsentation, die in der Praxis funktionieren – und das sind nicht Text- oder LaTeX-Dateien mit prädikatenlogischen Formeln; sicher nicht RDF/XML Quelltext; und auch nur sehr eingeschränkt Serialisierungen wie z. B. Turtle (Terse RDF Triple Language, ttl), auch wenn gerade Turtle explizit für eine menschliche Bearbeitung vorgesehen ist.

Wer meint, dass formale Ontologie wenig mit philosophischer Ontologie zu tun habe, irrt. Eine formale Ontologie modelliert nicht nur, bildet nicht nur ab, sondern formt und tradiert auch Sichtweisen auf Welt. Ontologien sind ein Teil der digitalen Realität, der allein durch die Formalisierungssprache deutliche Biases erzeugen kann. “An encoding bias results when representation choices are made purely for the convenience of notation or implementation.” ([Gr95], S. 3). Wo Ontologien mit digitalen Werkzeugen erstellt werden, wird das, was wir sehen, nicht nur durch die Ontologiesprache geformt, sondern auch durch die Visualisierungen und durch die Software, mit denen wir Ontologien erstellen (kommunizieren, nutzbar machen etc.).

6 Prä- vs. Post-Koordination

Die Bibliothekswissenschaft trägt mit einer grundlegenden Unterscheidung zu unserer Fragestellung bei. Die sogenannte *Prä-Koordination* setzt hier komplexe Klassen aus einfachen Klassen, Restriktionen etc. zusammen. Ziel ist es, ein klassifizierendes Objekt in möglichst genau eine möglichst genau passende und zu diesem Zweck möglichst genau beschriebene Klasse einzuordnen ([SS00], S.72). Ein Beispiel: Hengst = männliches Pferd, Rappe = schwarzes Pferd, Rapphengst = schwarzer Hengst (oder besser männlicher Rappe?). Prä-Koordination erzeugt im schlimmsten Fall eine kombinatorische Explosion von Klassen. Bei Mehrfach-Vererbungen greift eine Darstellung als Baum zu kurz, es entsteht ein gerichteter azyklischer Graph (DAG) von Subklassen-Beziehungen.

Im Gegensatz zur Prä-Koordination verzichtet die sogenannte *Post-Koordination* darauf, komplexe Klassen quasi auf Vorrat herzustellen. Stattdessen wird ein zu klassifizierendes Objekt in mehrere Klassen eingeordnet und damit als ein Punkt oder Unterraum in einem n-dimensionalen Merkmalsraum beschrieben. Wie kann man sich den Unterschied zwischen einer prä- und post-koordinierten Ontologie vorstellen? Hier sei eine Metapher erlaubt: Wer eine Prä-Koordination anstrebt, wird bei einer Ontologie an einen großen, fein verästelten Baum (oder allgemeiner ein DAG) aus Klassen denken. Wer eine Post-Koordination anstrebt, wird bei einer Ontologie an ein Wäldchen von kleinen, nur überschaubar verästelten Bäumchen denken – oder eben an einen höherdimensionalen Merkmalsraum.

Wenn man auf geschlossene, prä-koordinierende Begriffssysteme abzielt, lenkt man den Blick auf den Klassen-Baum oder den DAG. Wenn man auf ein post-koordinierendes Begriffssystem abzielt, sind weniger die Klassen an sich interessant als vielmehr deren Un-

terschiede. Ontologie lässt sich bei diesem Ansatz vor allem als eine Systematik von Unterschieden verstehen. Gewiss: Auch Unterschiede werden technisch mit Hilfe von Klassen kommuniziert. Aber es wechselt die Ordnung: Wer prä-koordinierend in einem Gegenstandsbereich Klassen sucht, konstruiert Klassen aus der „ersten Reihe“. Wer dagegen nach Unterschieden für eine Post-Koordination sucht, um Klassen aus der ersten Reihe unterscheiden zu können, konstruiert Klassen aus der „zweiten Reihe“.

7 Das Modellierungsbeispiel „Gewässer“

Wir wollen den Unterschied zwischen einer prä- und einer post-koordinierenden Ontologie an einem nichttrivialen Beispiel verdeutlichen. Ausgangsbasis ist der Inhalt der englischen Seite https://en.wikipedia.org/wiki/Body_of_water. Hier werden 91 Beispiele für Gewässer in Form eines Glossars aufgeführt: Verschiedenartigste Gewässer werden normalsprachlich anschaulich erläutert, wobei die Erläuterungen teilweise durch Links auf andere Wikipedia-Seiten angereichert sind. Bei den Erläuterungen der Einträge handelt es sich um frei formulierten Text, der von den Definitionen der verlinkten Wikipedia-Seiten unabhängig ist. Wir greifen diejenigen Beispiele heraus, die dann unten in Abb. 2 wieder eine Rolle spielen (Abb. 1):

Canal	an artificial waterway, usually connected to (and sometimes connecting) existing lakes, rivers, or oceans. (rc)
Channel	the physical confine of a river, slough or ocean strait consisting of a bed and banks. See also stream bed and strait. (rc)
Lagoon	a body of comparatively shallow salt or brackish water separated from the deeper sea by a shallow or exposed sandbank, coral reef, or similar feature. (nscm)
Lake	a body of water, usually freshwater, of relatively large size contained on a body of land. (nsc)
Maar	NO ENTRY (snc)
Puddle	a small accumulation of water on a surface, usually the ground. (tns)
Pond	a body of water smaller than a lake, especially those of artificial origin. (nsc)
Pool	various small bodies of water such as a swimming pool, reflecting pool, pond, or puddle. (nsc)
Reservoir	a place to store water for various uses, especially drinking water, which can be a natural or artificial (see lake and impoundment). (sc)
River	a natural waterway usually formed by water derived from either precipitation or glacial meltwater, and flows from higher ground to lower ground. (rnc)
Rivulet (UK, US literary)	a small or very small stream. (rnc)
Runnel	NO ENTRY (rnc)

Sea	a large expanse of saline water connected with an ocean, or a large, usually saline, lake that lacks a natural outlet such as the Caspian Sea and the Dead Sea. In common usage, often synonymous with the ocean. (nscm)
Stream	a body of water with a detectable current, confined within a bed and banks. (rc)
Tarn	a mountain lake or pool formed in a cirque excavated by a glacier. (nsc)
Torrent	NO ENTRY (rc)
Trickle	NO ENTRY (rc)

Abb. 1: Annotierter Textauszug aus der Seite https://en.wikipedia.org/wiki/Body_of_water

Zu den Annotationen in Abb. 1: Kürzel wie (rc) oder (nscm) etc. am Ende jedes Eintrags sind nicht auf obiger Seite enthalten, sondern stammen von mir und werden unten erklärt. Einträge mit NO ENTRY sind in obiger Seite nicht enthalten, sondern wurden von mir hinzugefügt, um weiter unter einen Vergleich mit der FCA zu ermöglichen.

Aus dieser Auswahl an Beispielen „im Kopf“ einen Baum (oder verallgemeinert einen DAG) aus Ober- und Unterklassen zu erzeugen, fällt schwer. Das liegt nicht an der Auswahl: Die Aufgabe wird nicht einfacher, wenn wir alle 91 Beispiele auf obiger Wikipedia-Seite betrachten. Wir suchen also einen systematischen Ansatz der Strukturierung. Eine mögliche Methode wäre diese:

1. Identifiziere für einen uns interessierenden Gegenstandsbereich (z. B. „Gewässer“) eine Menge von typischen und relevanten *Elementarsätzen*, z. B. obige Auswahl.
2. Tagge jeden Elementarsatz mit einem oder mehreren Elementen aus einer genau definierten Menge von Merkmalen. In unserem Beispiel könnten dies sein: *temporary, running, natural, stagnant, constant, maritime*.
3. Konstruiere aus der Menge von Merkmalsmengen mit einem systematischen Verfahren einen DAG aus Klassen-Inklusionen.

Schritt 1 liefert obige Wikipedia-Seite. Schritt 2 erzeugt die in Klammern notierten Buchstaben als Repräsentationen der Merkmale. Schritt 3 wird z. B. durch das Verfahren https://en.wikipedia.org/wiki/Formal_concept_analysis (FCA) geleistet. Das mathematische Ergebnis des FCA-Verfahrens ist eine halbgeordnete Menge (ein *Verband*), die man insbesondere als *Hasse-Diagramm* visualisieren kann, in unserem Fall wie in Abb. 2.

Das Wikipedia-Beispiel ist stark vereinfacht, zeigt aber das Prinzip: In der FCA taggen wir manuell oder automatisch einen Korpus von Elementarsätzen durch vorgegebene Merkmale und errechnen aus der entstehenden Tabelle ein Liniendiagramm. Das Liniendiagramm können wir nun sofort als Ontologie interpretieren oder darauf aufbauend die Auswahl der Merkmale bewerten, ggf. ändern, nach Bedarf verfeinern etc.

Wie man FCA als Werkzeug einsetzen kann – auch um Merkmale zu schöpfen –, zeigen Christoph Lübbert und Thomas Zeh anhand einer Abbildung aus dem Wikipedia-Artikel [https://de.wikipedia.org/wiki/Ontologie_\(Informatik\)](https://de.wikipedia.org/wiki/Ontologie_(Informatik)) kenntnisreich auf ([LZ22]).

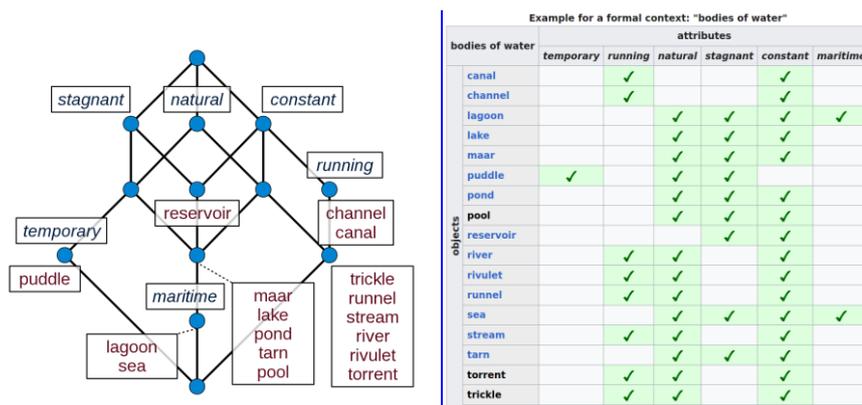


Abb. 2: Abbildungen aus https://en.wikipedia.org/wiki/Formal_concept_analysis. Rechts der sog. „formale Kontext“ mit den Beispielen aus Abb.1; links das daraus automatisch erzeugte Liniendiagramm.

Betrachtet man das Liniendiagramm aus Abb. 2 genauer, sieht man, dass in keinem Beispiel die Merkmale r und s gleichzeitig auftauchen: Zumindest in dem von uns analysierten Korpus aus 17 Beispielen gibt es offensichtlich keine Gewässer, die gleichzeitig fließend und stehend sind, und auch keine, die gleichzeitig fließend und temporär sind.

Lebensweltliches Wissen bestätigt: Fließend oder stehend schließt sich sachlogisch aus. Fließend und temporär hingegen erzeugt kein Problem– hier fehlen uns in obiger Auswahl einfach noch Beispiele, die diese Kombination zeigen. In obiger FCA werden die Merkmale als unstrukturierte Menge betrachtet. Es wäre aber durchaus wünschenswert, auch bezüglich der Merkmale etwas Hintergrundwissen erfassen zu können, also die Merkmale in Teilmengen von sachlogisch zusammengehörigen (und einander genau deshalb ausschließenden) Merkmalen zu ordnen. Die FCA erlaubt es, dies technisch zu tun, lediglich das Wikipedia-Beispiel zur FCA ist hier nicht mehr differenziert genug. Ergänzend zur technischen Machbarkeit stellt sich die Frage, woher dieses Hintergrundwissen kommt: Wird es der FCA durch eine geeignete Datenstruktur ex ante mitgegeben, oder lässt es sich ex post – ggf. auch iterativ – aus einer Iteration aus FCA-Durchläufen ablesen?

Wir sehen: Eine (Primär-)Systematik von Gewässern lässt sich auf einer (Sekundär-)Systematik jener Merkmale aufbauen, mit denen man Gewässer unterscheiden kann. Eine mögliche Grundlage der Sekundär-Systematik sind Gruppierungen wie z. B. stehend vs. fließend, temporär vs. dauerhaft, Süß- vs. Salzwasser usw.

Die deutsche Seite <https://de.wikipedia.org/wiki/Gewässer> macht genau das: Sie entwickelt explizit und hoch strukturiert, wonach sich Gewässer unterscheiden lassen, schon im Inhaltsverzeichnis:

1.2 Nach Lage des Wasserkörpers [JB: oberirdisch, unterirdisch]

1.3 Nach Stellung im Gewässersystem [JB: Binnengewässer, Meere]

1.4 Nach dem Strömungsverhalten [JB: strömend, stehend]

1.5 Nach Wasserführung im Zeitverlauf [JB: perennierend, intermittierend]

Diese Wikipedia-Seite klassifiziert nicht Gewässer selbst, sondern sie entwickelt Klassifikationskriterien von Gewässern. Wir erkennen hier unschwer das Prinzip der Post-Koordination wieder.

Wer – ausgehend von einer Aufzählung von Beispielen über eine Systematik von Merkmalen hin zu einer Ontologie – einen komplexen Gegenstandsbereich begrifflich systematisch ausdifferenzieren und formal darstellen will, ist zunehmend an einer Systematik von Unterschieden interessiert. Interessant sind auf diesem Weg weniger die prä-koordinierten Klassen an sich als vielmehr die Merkmale und ihre Zusammenhänge, aus denen sich solche Klassen dann bei Bedarf kombinatorisch herstellen lassen. Nicht alle Kombinationen von Merkmalen sind in der Praxis relevant. So werden in der Praxis z. B. nur Binnengewässer, aber keine Meere nach ihrem Strömungsverhalten unterschieden; eine Gewässerordnungszahl wird nur für Flüsse angegeben, usw. Auch solche Informationen sollten in einer Systematik der Unterschiede enthalten sein.

Eine Eingangsfrage dieses Aufsatzes lautet: *Was stelle ich als Dozent mir vor, wenn ich den Studierenden zeigen will, was eine Ontologie „ist“?* Auf Grundlage obiger Argumentation sind wir bei folgender Position angekommen:

- Weniger interessant als ein Klassen-Baum (oder Netz, DAG, Hasse-Diagramm etc.) ist eine Systematik von möglichen Unterschieden zwischen Klassen.
- Eine Ontologie zu bauen, heißt darzustellen, welche Unterschiede in einem bestimmten Praxiskontext tatsächlich einen Unterschied machen.
- *Es macht einen Unterschied, ob man Gemeinsamkeiten oder Unterschiede systematisiert.* ([Zi00]).

Wenn die Visualisierung bestimmt, was man sieht; und wenn unsere Werkzeuge bestimmen, was wir gestalten können; und wenn die Grenzen unserer Sprache die Grenzen unserer Welt sind (Wittgenstein): Dann wollen wir im Kontext unserer Fragestellung nach einer Visualisierung, einem Tool, einer Sprache suchen, mit der wir Unterschiede systematisieren können.

Im vergangenen Jahr wurde u. a. im AKWI-Tagungsband die Sprache GenDifS vorgestellt, mit der sich in dem Open-Source-Tool Freemind Taxonomien grafisch entwickeln und über ein Python-Skript nach OWL und SKOS exportieren lassen ([Bu22]. GenDifS

stellt eine konzise Syntax bereit, mit der sich insbesondere ein Pattern realisieren lässt, das z. B. in <https://www.w3.org/TR/2005/NOTE-swbp-specified-values-20050517/> beschrieben ist. GenDifS realisiert das seit Aristoteles bekannte Pattern *genus proximum, differentia specifica*. Im Unterschied zu anderen editierbaren Ontologie-Visualisierungen werden in GenDifS Klassen und ihre charakteristischen Unterschiede (hier Attribut-Wert-Paare) in direkter Nachbarschaft visuell modelliert.

8 Gewässer mit GenDifS

GenDifS scheint besonders geeignet zu sein, die verschiedenen Facetten einer Post-Koordination zu modellieren. Wir probieren das am Gewässer-Beispiel aus. Es entsteht folgende Mindmap:



Abb. 3: Wikipedia > Gewässer, Struktur in GenDifS

Diese Mindmap lässt sich automatisch in eine SKOS-Terminologie und eine OWL-Ontologie übersetzen, womit diese spezielle Mindmap eine formale Semantik besitzt. Die genaue formale Semantik dieser Beispiel-Mindmap ist in der Online-Fassung dieses Aufsatzes unter <http://jbusse.de/akwi2023/> > [Semantik Gewässer-Beispiel](#) beschrieben. Eine Übersicht über die formale Semantik aller GenDifS-Sprachelemente gibt die Seite <http://www.jbusse.de/gendifs/> > [GenDifS Formal Semantics](#). Es scheint, dass hier recht explizit klar wird, an welcher Stelle Unterschiede einen Unterschied machen – mehr als in anderen uns bekannten Visualisierungen einer Ontologie.

Exkurs. Auch im vorliegenden Aufsatz suchen wir ja offensichtlich nach Unterschieden, hier auf der Ebene von Modell-Typen. Was unterscheidet ein GenDifS-Modell von anderen Modelltypen? Das ließe sich durch einen Vergleich des GenDifS-Modells zum Gewässerbeispiel mit anderen Modellen und Visualisierungen zum selben Beispiel analysieren. Ergänzend wäre dann allerdings auch ein Vergleich der formalen Semantik der verschiedenen Modelle erforderlich: „Stimmt“ das denn, was da intuitiv als – ggf. wie in Protégé durch OWL-Code ergänzbare – Visualisierung modelliert wurde? Was wird tatsächlich ausgedrückt?

GenDifS implementiert das oben zitierte Pattern *Values as subclasses partitioning a “feature”* und erzeugt damit einen vergleichsweise komplexen OWL-Code, wie er manuell kaum fehlerfrei codiert werden könnte. Es stellt sich kritisch die Frage nach dem Mehrwert dieser Komplexität: Wie wird das erzeugte Modell von einer bestimmten Applikation für

einen bestimmten Usecase in einem bestimmten praxisorientierten Zusammenhang ausgewertet? Andererseits scheint GenDifS auch Komplexität zu reduzieren, indem es für ein einziges, für Taxonomien jedoch zentrales Pattern einen niederschweligen, visualisierten Zugang auch für Nichtexperten eröffnet. Darin liegt u. E. die Leistung dieses speziellen Tools insbesondere für die Erstellung von post-koordinierenden Begriffssystemen. Ende des Exkurses.

9 Zusammenfassung

Das im Gewässer-Beispiel entwickelte GenDifS-Modell visualisiert nicht nur eine Menge von möglichen Unterschieden, sondern hat auch eine formale Semantik. (Die Website <http://jbusse.de/akwi2023/> geht hierauf im Detail ein.) Im Wesentlichen werden Klassen erzeugt, die es erlauben, Dinge anhand ihrer charakteristischen Merkmale in Subclasses einzuordnen. In unserem Gewässer-Beispiel: Es gibt Meere und Binnengewässer; Binnengewässer – und nur diese – lassen sich sinnvoll in fließende und stehende Gewässer unterscheiden; ein Binnengewässer, das fließt, ist ein Fließgewässer; usw. Solche Systematiken zu modellieren, ist auch direkt in Protégé und OWL möglich, aber deutlich komplexer und fehleranfälliger.

Neu an GenDifS ist: (1) GenDifS-Taxonomien werden aus einer editierbaren Visualisierung generiert, die auch Studierenden ohne spezifische Ausbildung in OWL zugänglich ist; (2) die spezifische Verbindung von Visualisierung und Sprachelementen unterstützt es, eine Vorstellung von Ontologien zu entwickeln, in der eine Domäne nicht prä-koordinierend als Baum (oder DAG oder auch FCA), sondern post-koordinierend als eine Systematik von Merkmalen geordnet wird, die für Unterscheidungen herangezogen werden können.

Aus mathematischer Sicht bedeutet der Wechsel von Prä- nach Post-Koordination, dass man einen Gegenstandsbereich nicht mehr primär als eine Systematik (Baum, DAG etc.) von Klassen modelliert, sondern dass man nach einem n-dimensionalen Merkmalsraum sucht, der einen Gegenstandsbereich in Teilräume untergliedert. In der Praxis führt der Wechsel von „prä“ nach „post“ dazu, dass nicht mehr der Entwickler einer Ontologie schon alle Eventualitäten vorhersehen muss. Stattdessen kann die Entscheidung, welche Merkmale in einer praktischen Aufgabenstellung für eine Klassenbildung relevant werden, ad hoc getroffen werden: in der Praxis, in einer konkreten Situation, in einer bestimmten Rolle, aus einer bestimmten Perspektive – mit all den subjektiven Zutaten also, die eine Problemlösung in der DSR beeinflussen.

10 URL-Verzeichnis

In diesem Aufsatz sind verschiedene Seiten aus dem Web primärer Gegenstand der Betrachtung. Wir halten es im Kontext des Semantic Web für sinnvoll, solche Seiten als *first class citizens* zu behandeln und auch im Text über ihre URI zu identifizieren, falls diese eine sinntragende und für Kommunikation hilfreiche Repräsentation besitzt. Wir fassen diese Links deshalb in einem eigenen URL-Verzeichnis zusammen.

https://en.wikipedia.org/wiki/Body_of_water A body of water [...] is any significant accumulation of water [...]; (2023-03-10)

<https://de.wikipedia.org/wiki/Gewässer> Ein Gewässer ist in der Natur fließendes oder stehendes Wasser. (2023-06-08)

[https://de.wikipedia.org/wiki/Ontologie_\(Informatik\)](https://de.wikipedia.org/wiki/Ontologie_(Informatik)) Ontologie (Informatik) (2023-03-19)

https://en.wikipedia.org/wiki/Formal_concept_analysis In information science, formal concept analysis (FCA) is a principled way of deriving a concept hierarchy or formal ontology from a collection of objects and their properties.

<http://jbusse.de/akwi2023/> Online-Version und Begleitmaterial zum Aufsatz J.Busse: *Prä- versus postkoordinierende Ontologien*, AKWI 2023

<http://www.jbusse.de/gendifs/> GenDifS 0.6 (2023-06-20)

<https://protege.stanford.edu/> A free, open-source ontology editor and framework for building intelligent systems. Aktive Version im Feb 2023: 5.2.3

<https://www.govdata.de/web/guest/sparql-assistent> Der auf GovData veröffentlichte Metadatenkatalog lässt sich [...] auch maschinell abfragen und verarbeiten. Hierfür stehen ein SPARQL Endpunkt sowie die CKAN API (<https://www.govdata.de/ckan/api>) zur Verfügung.

<https://www.govdata.de> GovData ist das Datenportal für Deutschland. Im Informationsbereich finden Sie alles Wissenswerte rund um die Themen OpenData, OpenGovernment und Bürgerbeteiligung, ...

<https://www.w3.org/2016/05/ontolex/#core> Figure 1 Lemon OntoLex Core.png, aus [CMB16]

<https://www.w3.org/TR/2005/NOTE-swbp-specified-values-20050517/> siehe [Re05]

11 Literatur

[Bi21] Biagetti, Maria Teresa. 2021. "Ontologies as knowledge organization systems". Knowledge Organization 48, no. 2: 152-176. Also available in ISKO Encyclopedia of Knowledge Organization, eds. Birger Hjørland and Claudio Gnoli, <https://www.isko.org/cyclo/ontologies>

[Bu22] Busse, J: Kernkonzepte der Taxonomiesprache GenDifS, in: Eggert, S (Hrsg.): Tagungsband zur 35. Jahrestagung des AKWI, 2022. https://doi.org/10.30844/AKWI_2022_14.

- [Bu23] Busse, J.; Reibold, A. (Organizers): Ontologie, Linguistik, Terminologie, Logik. Forschung trifft Praxis. Dagstuhl Research Meeting 23144, Apr 02 – Apr 05, 2023. <https://www.dagstuhl.de/seminars/seminar-calendar/seminar-details/23144>
- [CMB16] Cimiano, P.; McCrae, J.; Buitelaar, P.: Lexicon Model for Ontologies. Community Report, 10 May 2016. Final Community Group Report 10 May 2016 <https://www.w3.org/2016/05/ontolex/>
- [DCAT22] AG für GovData: [DCAT-AP.de](https://www.dcat-ap.de) Spezifikation 2.0: Deutsche Adaption des „Data Catalogue Application Profile“ (DCAT-AP) für Datenportale in Europa, 1. März 2022, zuletzt geändert am 28. Februar 2022. <https://www.dcat-ap.de/def/dcatde/2.0/spec/specification.pdf>
- [HMP+04] Hevner, A.; March, S.; Park, J. et al: Design Science in Information Systems Research. MIS Quarterly Vol. 28 No. 1, pp. 75-105/March 2004. https://www.researchgate.net/publication/201168946_Design_Science_in_Information_Systems_Research
- [Gr95] Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing? International Journal of Human-Computer Studies, 43(5-6):907–928, November 1995. <https://doi.org/10.1006/ijhc.1995.1081>
- [LZ22] Lübbert, C.; Zeh, T.: Skizze eines Verfahrens zur Erstellung von Ontologien mittels Formaler Begriffsanalyse. Informatik Spektrum (2022) 45:3–12, <https://doi.org/10.1007/s00287-021-01397-1>
- [Mu15] Musen, M.A. The Protégé project: A look back and a look forward. AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4), June 2015. DOI: 10.1145/2557001.25757003.
- [Re05] Rector, A.: Representing Specified Values in OWL: “value partitions” and “value sets”. W3C Working Group Note 17 May 2005. <https://www.w3.org/TR/swbp-specified-values/>
- [SS00] Stock, W.; Stock, M: Wissensrepräsentation. Informationen auswerten und bereitstellen. Oldenbourg 2008. <https://doi.org/10.1524/9783486844900>
- [St11] Suckenschmidt, H.: Ontologien. Springer, 2. Auflage 2011
- [Zi00] Zimmerli, W. Ch.: Vom Unterschied, der einen Unterschied macht. Information, Netzwerkdenken und Mensch-Maschine-Tandem. Aus: Mittelstraß, Jürgen (Hrsg.): Die Zukunft des Wissens. Akademie Verlag, 2000, S.115-128. <https://doi.org/10.1515/9783050078618-014>

Dieser Aufsatz liegt auch online unter <http://jbusse.de/akwi2023/> in einer html-Fassung vor. Dort finden sich auch ergänzende Materialien.